REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-

0777

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE December, 1998	3. R

4. TITLE AND SUBTITLE 1998 Summer Research Program (SRP), Summer Faculty Research Program (SFRP), Final Reports, Volume 2, Armstrong Laboratory	5. FUNDING NUMBERS F49620-93-C-0063
6. AUTHOR(S) Gary Moore	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Research & Development Laboratories (RDL) 5800 Uplander Way Culver City, CA 90230-6608	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research (AFOSR) 801 N. Randolph St. Arlington, VA 22203-1977	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTARY NOTES	
12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release	12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The United States Air Force Summer Research Program (USAF-SRP) is designed to introduce university, college, and technical institute faculty members, graduate students, and high school students to Air Force research. This is accomplished by the faculty members (Summer Faculty Research Program, (SFRP)), graduate students (Graduate Student Research Program (GSRP)), and high school students (High School Apprenticeship Program (HSAP)) being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory (AFRL) Technical Directorates, Air Force Air Logistics Centers (ALC), and other AF Laboratories. This volume consists of a program overview, program management statistics, and the final technical reports from the SFRP participants at the Armstrong Laboratory.

14. SUBJECT TERMS Air Force Research, Air Force, Engineering, Laboratories, Reports, Summer, Universities, Faculty, Graduate Student, High School Student	15. NUMBER OF PAGES
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements.***

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C - Contract
G - Grant
PE - Program Element
PR - Project
TA - Task
WU - Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with....; Trans. of....; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.
DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
Leave blank.
NASA - Leave blank.
NTIS -

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self- explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1998

SUMMER FACULTY RESEARCH PROGRAM FINAL REPORTS

VOLUME 2

ARMSTRONG LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Colonel Jan Cerveny

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Johnetta Thompson

Program Administrator, RDL
Rebecca Kelly-Clemmons

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1998

PREFACE

Reports in this volume are numbered consecutively beginning with number 1. Each report is paginated with the report number followed by consecutive page numbers, e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3.

This document is one of a set of 15 volumes describing the 1998 AFOSR Summer Research Program. The following volumes comprise the set:

VOLUME	TITLE
1	Program Management Report
	Summer Faculty Research Program (SFRP) Reports
2	Armstrong Laboratory
3	Phillips Laboratory
4	Rome Laboratory
5A & 5B	Wright Laboratory
6	Arnold Engineering Development Center, Air Logistics Centers, United States Air Force Academy and Wilford Hall Medical Center
	Graduate Student Research Program (GSRP) Reports
7	Armstrong Laboratory
8	Phillips Laboratory
9	Rome Laboratory
10	Wright Laboratory
11	Arnold Engineering Development Center, and Wilford Hall Medical Center
	High School Apprenticeship Program (HSAP) Reports
12	Armstrong Laboratory
13	Phillips Laboratory
14	Rome Laboratory
15A, 15B & 15C	Wright Laboratory

SFRP FINAL REPORT TABLE OF CONTENTS i-x

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

1. INTRODUCTION

The Summer Research Program (SRP), sponsored by the Air Force Office of Scientific Research (AFOSR), offers paid opportunities for university faculty, graduate students, and high school students to conduct research in U.S. Air Force research laboratories nationwide during the summer.

Introduced by AFOSR in 1978, this innovative program is based on the concept of teaming academic researchers with Air Force scientists in the same disciplines using laboratory facilities and equipment not often available at associates' institutions.

The Summer Faculty Research Program (SFRP) is open annually to approximately 150 faculty members with at least two years of teaching and/or research experience in accredited U.S. colleges, universities, or technical institutions. SFRP associates must be either U.S. citizens or permanent residents.

The Graduate Student Research Program (GSRP) is open annually to approximately 100 graduate students holding a bachelor's or a master's degree; GSRP associates must be U.S. citizens enrolled full time at an accredited institution.

The High School Apprentice Program (HSAP) annually selects about 125 high school students located within a twenty mile commuting distance of participating Air Force laboratories.

AFOSR also offers its research associates an opportunity, under the Summer Research Extension Program (SREP), to continue their AFOSR-sponsored research at their home institutions through the award of research grants. In 1994 the maximum amount of each grant was increased from $20,000 to $25,000, and the number of AFOSR-sponsored grants decreased from 75 to 60. A separate annual report is compiled on the SREP.

The numbers of projected summer research participants in each of the three categories and SREP "grants" are usually increased through direct sponsorship by participating laboratories.

AFOSR's SRP has well served its objectives of building critical links between Air Force research laboratories and the academic community, opening avenues of communications and forging new research relationships between Air Force and academic technical experts in areas of national interest, and strengthening the nation's efforts to sustain careers in science and engineering. The success of the SRP can be gauged from its growth from inception (see Table 1) and from the favorable responses the 1997 participants expressed in end-of-tour SRP evaluations (Appendix B).

AFOSR contracts for administration of the SRP by civilian contractors. The contract was first awarded to Research & Development Laboratories (RDL) in September 1990. After completion of the 1990 contract, RDL (in 1993) won the recompetition for the basic year and four 1-year options.

2. PARTICIPATION IN THE SUMMER RESEARCH PROGRAM

The SRP began with faculty associates in 1979; graduate students were added in 1982 and high school students in 1986. The following table shows the number of associates in the program each year.

YEAR	SRP Participation, by Year			TOTAL
	SFRP	GSRP	HSAP	
1979	70			70
1980	87			87
1981	87			87
1982	91	17		108
1983	101	53		154
1984	152	84		236
1985	154	92		246
1986	158	100	42	300
1987	159	101	73	333
1988	153	107	101	361
1989	168	102	103	373
1990	165	121	132	418
1991	170	142	132	444
1992	185	121	159	464
1993	187	117	136	440
1994	192	117	133	442
1995	190	115	137	442
1996	188	109	138	435
1997	148	98	140	427
1998	85	40	88	213

Beginning in 1993, due to budget cuts, some of the laboratories weren't able to afford to fund as many associates as in previous years. Since then, the number of funded positions has remained fairly constant at a slightly lower level.

3. RECRUITING AND SELECTION

The SRP is conducted on a nationally advertised and competitive-selection basis. The advertising for faculty and graduate students consisted primarily of the mailing of 8,000 52-page SRP brochures to chairpersons of departments relevant to AFOSR research and to administrators of grants in accredited universities, colleges, and technical institutions. Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs) were included. Brochures also went to all participating USAF laboratories, the previous year's participants, and numerous individual requesters (over 1000 annually).

RDL placed advertisements in the following publications: *Black Issues in Higher Education, Winds of Change,* and *IEEE Spectrum.* Because no participants list either *Physics Today* or *Chemical & Engineering News* as being their source of learning about the program for the past several years, advertisements in these magazines were dropped, and the funds were used to cover increases in brochure printing costs.

High school applicants can participate only in laboratories located no more than 20 miles from their residence. Tailored brochures on the HSAP were sent to the head counselors of 180 high schools in the vicinity of participating laboratories, with instructions for publicizing the program in their schools. High school students selected to serve at Wright Laboratory's Armament Directorate (Eglin Air Force Base, Florida) serve eleven weeks as opposed to the eight weeks normally worked by high school students at all other participating laboratories.

Each SFRP or GSRP applicant is given a first, second, and third choice of laboratory. High school students who have more than one laboratory or directorate near their homes are also given first, second, and third choices.

Laboratories make their selections and prioritize their nominees. AFOSR then determines the number to be funded at each laboratory and approves laboratories' selections.

Subsequently, laboratories use their own funds to sponsor additional candidates. Some selectees do not accept the appointment, so alternate candidates are chosen. This multi-step selection procedure results in some candidates being notified of their acceptance after scheduled deadlines. The total applicants and participants for 1998 are shown in this table.

1998 Applicants and Participants			
PARTICIPANT CATEGORY	TOTAL APPLICANTS	SELECTEES	DECLINING SELECTEES
SFRP	382	85	13
(HBCU/MI)	(0)	(0)	(0)
GSRP	130	40	7
(HBCU/MI)	(0)	(0)	(0)
HSAP	328	88	22
TOTAL	840	213	42

4. SITE VISITS

During June and July of 1998, representatives of both AFOSR/NI and RDL visited each participating laboratory to provide briefings, answer questions, and resolve problems for both laboratory personnel and participants. The objective was to ensure that the SRP would be as constructive as possible for all participants. Both SRP participants and RDL representatives found these visits beneficial. At many of the laboratories, this was the only opportunity for all participants to meet at one time to share their experiences and exchange ideas.

5. HISTORICALLY BLACK COLLEGES AND UNIVERSITIES AND MINORITY INSTITUTIONS (HBCU/MIs)

Before 1993, an RDL program representative visited from seven to ten different HBCU/MIs annually to promote interest in the SRP among the faculty and graduate students. These efforts were marginally effective, yielding a doubling of HBCI/MI applicants. In an effort to achieve AFOSR's goal of 10% of all applicants and selectees being HBCU/MI qualified, the RDL team decided to try other avenues of approach to increase the number of qualified applicants. Through the combined efforts of the AFOSR Program Office at Bolling AFB and RDL, two very active minority groups were found, HACU (Hispanic American Colleges and Universities) and AISES (American Indian Science and Engineering Society). RDL is in communication with representatives of each of these organizations on a monthly basis to keep up with the their activities and special events. Both organizations have widely-distributed magazines/quarterlies in which RDL placed ads.

Since 1994 the number of both SFRP and GSRP HBCU/MI applicants and participants has increased ten-fold, from about two dozen SFRP applicants and a half dozen selectees to over 100 applicants and two dozen selectees, and a half-dozen GSRP applicants and two or three selectees to 18 applicants and 7 or 8 selectees. Since 1993, the SFRP had a two-fold applicant increase and a two-fold selectee increase. Since 1993, the GSRP had a three-fold applicant increase and a three to four-fold increase in selectees.

In addition to RDL's special recruiting efforts, AFOSR attempts each year to obtain additional funding or use leftover funding from cancellations the past year to fund HBCU/MI associates.

SRP HBCU/MI Participation, By Year				
YEAR	SFRP		GSRP	
	Applicants	Participants	Applicants	Participants
1985	76	23	15	11
1986	70	18	20	10
1987	82	32	32	10
1988	53	17	23	14
1989	39	15	13	4
1990	43	14	17	3
1991	42	13	8	5
1992	70	13	9	5
1993	60	13	6	2
1994	90	16	11	6
1995	90	21	20	8
1996	119	27	18	7

6. SRP FUNDING SOURCES

Funding sources for the 1998 SRP were the AFOSR-provided slots for the basic contract and laboratory funds. Funding sources by category for the 1998 SRP selected participants are shown here.

1998 SRP FUNDING CATEGORY	SFRP	GSRP	HSAP
AFOSR Basic Allocation Funds	67	38	75
USAF Laboratory Funds	17	2	13
Slots Added by AFOSR (Leftover Funds)	0	0	0
HBCU/MI By AFOSR (Using Procured Addn'l Funds)	0	0	N/A
TOTAL	84	40	88

7. COMPENSATION FOR PARTICIPANTS

Compensation for SRP participants, per five-day work week, is shown in this table.

1998 SRP Associate Compensation

PARTICIPANT CATEGORY	1991	1992	1993	1994	1995	1996	1997	1998
Faculty Members	$690	$718	$740	$740	$740	$770	$770	$793
Graduate Student (Master's Degree)	$425	$442	$455	$455	$455	$470	$470	$484
Graduate Student (Bachelor's Degree)	$365	$380	$391	$391	$391	$400	$400	$412
High School Student (First Year)	$200	$200	$200	$200	$200	$200	$200	$200
High School Student (Subsequent Years)	$240	$240	$240	$240	$240	$240	$240	$240

The program also offered associates whose homes were more than 50 miles from the laboratory an expense allowance (seven days per week) of $52/day for faculty and $41/day for graduate students. Transportation to the laboratory at the beginning of their tour and back to their home destinations at the end was also reimbursed for these participants. Of the combined SFRP and GSRP associates, 65 % claimed travel reimbursements at an average round-trip cost of $730.

Faculty members were encouraged to visit their laboratories before their summer tour began. All costs of these orientation visits were reimbursed. Forty-three percent (85 out of 188) of faculty associates took orientation trips at an average cost of $449. By contrast, in 1993, 58 % of SFRP associates elected to take an orientation visits at an average cost of $685; that was the highest percentage of

associates opting to take an orientation trip since RDL has administered the SRP, and the highest average cost of an orientation trip.

Program participants submitted biweekly vouchers countersigned by their laboratory research focal point, and RDL issued paychecks so as to arrive in associates' hands two weeks later.

This is the third year of using direct deposit for the SFRP and GSRP associates. The process went much more smoothly with respect to obtaining required information from the associates, about 15% of the associates' information needed clarification in order for direct deposit to properly function as opposed to 7% from last year. The remaining associates received their stipend and expense payments via checks sent in the US mail.

HSAP program participants were considered actual RDL employees, and their respective state and federal income tax and Social Security were withheld from their paychecks. By the nature of their independent research. SFRP and GSRP program participants were considered to be consultants or independent contractors. As such, SFRP and GSRP associates were responsible for their own income taxes, Social Security. and insurance.

8. CONTENTS OF THE 1998 REPORT

The complete set of reports for the 1998 SRP includes this program management report (Volume 1) augmented by fifteen volumes of final research reports by the 1998 associates, as indicated below:

1998 SRP Final Report Volume Assignments

LABORATORY	SFRP	GSRP	HSAP
Armstrong	2	7	12
Phillips	3	8	13
Rome	4	9	14
Wright	5A, 5B	10	15
AEDC, ALCs, USAFA, WHMC	6	11	

APPENDIX A -- PROGRAM STATISTICAL SUMMARY

A. Colleges/Universities Represented

Selected SFRP associates represented 169 different colleges, universities, and institutions, GSRP associates represented 95 different colleges, universities, and institutions.

B. States Represented

SFRP -Applicants came from 47 states plus Washington D.C. Selectees represent 44 states.

GSRP - Applicants came from 44 states. Selectees represent 32 states.

HSAP - Applicants came from thirteen states. Selectees represent nine states.

Total Number of Participants	
SFRP	85
GSRP	40
HSAP	88
TOTAL	213

Degrees Represented			
	SFRP	GSRP	TOTAL
Doctoral	83	0	83
Master's	1	3	4
Bachelor's	0	22	22
TOTAL	186	25	109

SFRP Academic Titles	
Assistant Professor	36
Associate Professor	34
Professor	15
Instructor	0
Chairman	0
Visiting Professor	0
Visiting Assoc. Prof.	0
Research Associate	0
TOTAL	85

Source of Learning About the SRP		
Category	Applicants	Selectees
Applied/participated in prior years	177	47
Colleague familiar with SRP	104	24
Brochure mailed to institution	101	21
Contact with Air Force laboratory	101	39
IEEE Spectrum	12	1
BIIHE	4	0
Other source	117	30
TOTAL	616	162

APPENDIX B -- SRP EVALUATION RESPONSES

1. OVERVIEW

Evaluations were completed and returned to RDL by four groups at the completion of the SRP. The number of respondents in each group is shown below.

Table B-1. Total SRP Evaluations Received

Evaluation Group	Responses
SFRP & GSRPs	100
HSAPs	75
USAF Laboratory Focal Points	84
USAF Laboratory HSAP Mentors	6

All groups indicate unanimous enthusiasm for the SRP experience.

The summarized recommendations for program improvement from both associates and laboratory personnel are listed below:

A. Better preparation on the labs' part prior to associates' arrival (i.e., office space, computer assets, clearly defined scope of work).

B. Faculty Associates suggest higher stipends for SFRP associates.

C. Both HSAP Air Force laboratory mentors and associates would like the summer tour extended from the current 8 weeks to either 10 or 11 weeks; the groups state it takes 4-6 weeks just to get high school students up-to-speed on what's going on at laboratory. (Note: this same argument was used to raise the faculty and graduate student participation time a few years ago.)

2. 1998 USAF LABORATORY FOCAL POINT (LFP) EVALUATION RESPONSES

The summarized results listed below are from the 84 LFP evaluations received.

1. LFP evaluations received and associate preferences:

Table B-2. Air Force LFP Evaluation Responses (By Type)

		How Many Associates Would You Prefer To Get ? (% Response)											
		SFRP				GSRP (w/Univ Professor)				GSRP (w/o Univ Professor)			
Lab	**Evals Recv'd**	**0**	**1**	**2**	**3+**	**0**	**1**	**2**	**3+**	**0**	**1**	**2**	**3+**
AEDC	0	-	-	-	-	-	-	-	-	-	-	-	-
WHMC	0	-	-	-	-	-	-	-	-	-	-	-	-
AL	7	28	28	28	14	54	14	28	0	86	0	14	0
USAFA	1	0	100	0	0	100	0	0	0	0	100	0	0
PL	25	40	40	16	4	88	12	0	0	84	12	4	0
RL	5	60	40	0	0	80	10	0	0	100	0	0	0
WL	46	30	43	20	6	78	17	4	0	93	4	2	0
Total	**84**	**32%**	**50%**	**13%**	**5%**	**80%**	**11%**	**6%**	**0%**	**73%**	**23%**	**4%**	**0%**

LFP Evaluation Summary. The summarized responses, by laboratory, are listed on the following page. LFPs were asked to rate the following questions on a scale from 1 (below average) to 5 (above average).

2. LFPs involved in SRP associate application evaluation process:
   a. Time available for evaluation of applications:
   b. Adequacy of applications for selection process:
3. Value of orientation trips:
4. Length of research tour:
5 a. Benefits of associate's work to laboratory:
   b. Benefits of associate's work to Air Force:
6. a. Enhancement of research qualifications for LFP and staff:
   b. Enhancement of research qualifications for SFRP associate:
   c. Enhancement of research qualifications for GSRP associate:
7. a. Enhancement of knowledge for LFP and staff:
   b. Enhancement of knowledge for SFRP associate:
   c. Enhancement of knowledge for GSRP associate:
8. Value of Air Force and university links:
9. Potential for future collaboration:
10. a. Your working relationship with SFRP:
    b. Your working relationship with GSRP:
11. Expenditure of your time worthwhile:

(Continued on next page)

12. Quality of program literature for associate:
13. a. Quality of RDL's communications with you:
    b. Quality of RDL's communications with associates:
14. Overall assessment of SRP:

Table B-3. Laboratory Focal Point Reponses to above questions

	AEDC	*AL*	*USAFA*	*PL*	*RL*	*WHMC*	*WL*
# Evals Recv'd	0	7	1	14	5	0	46
Question #							
2	-	86 %	0 %	88 %	80 %	-	85 %
2a	-	4.3	n/a	3.8	4.0	-	3.6
2b	-	4.0	n/a	3.9	4.5	-	4.1
3	-	4.5	n/a	4.3	4.3	-	3.7
4	-	4.1	4.0	4.1	4.2	-	3.9
5a	-	4.3	5.0	4.3	4.6	-	4.4
5b	-	4.5	n/a	4.2	4.6	-	4.3
6a	-	4.5	5.0	4.0	4.4	-	4.3
6b	-	4.3	n/a	4.1	5.0	-	4.4
6c	-	3.7	5.0	3.5	5.0	-	4.3
7a	-	4.7	5.0	4.0	4.4	-	4.3
7b	-	4.3	n/a	4.2	5.0	-	4.4
7c	-	4.0	5.0	3.9	5.0	-	4.3
8	-	4.6	4.0	4.5	4.6	-	4.3
9	-	4.9	5.0	4.4	4.8	-	4.2
10a	-	5.0	n/a	4.6	4.6	-	4.6
10b	-	4.7	5.0	3.9	5.0	-	4.4
11	-	4.6	5.0	4.4	4.8	-	4.4
12	-	4.0	4.0	4.0	4.2	-	3.8
13a	-	3.2	4.0	3.5	3.8	-	3.4
13b	-	3.4	4.0	3.6	4.5	-	3.6
14	-	4.4	5.0	4.4	4.8	-	4.4

3. 1998 SFRP & GSRP EVALUATION RESPONSES

The summarized results listed below are from the 120 SFRP/GSRP evaluations received.

Associates were asked to rate the following questions on a scale from 1 (below average) to 5 (above average) - by Air Force base results and over-all results of the 1998 evaluations are listed after the questions.

1. The match between the laboratories research and your field:
2. Your working relationship with your LFP:
3. Enhancement of your academic qualifications:
4. Enhancement of your research qualifications:
5. Lab readiness for you: LFP, task, plan:
6. Lab readiness for you: equipment, supplies, facilities:
7. Lab resources:
8. Lab research and administrative support:
9. Adequacy of brochure and associate handbook:
10. RDL communications with you:
11. Overall payment procedures:
12. Overall assessment of the SRP:
13. a. Would you apply again?
    b. Will you continue this or related research?
14. Was length of your tour satisfactory?
15. Percentage of associates who experienced difficulties in finding housing:
16. Where did you stay during your SRP tour?
    a. At Home:
    b. With Friend:
    c. On Local Economy:
    d. Base Quarters:
17. Value of orientation visit:
    a. Essential:
    b. Convenient:
    c. Not Worth Cost:
    d. Not Used:

SFRP and GSRP associate's responses are listed in tabular format on the following page.

Table B-4. 1997 SFRP & GSRP Associate Responses to SRP Evaluation

# res	Arnold 6	Brooks 48	Edwards 6	Eglin 14	Griffis 31	Hanscom 19	Kelly 3	Kirtland 32	Lackland 1	Robins 2	Tyndall 10	WPAFB 85	average 257
1	4.8	4.4	4.6	4.7	4.4	4.9	4.6	4.6	5.0	5.0	4.0	4.7	4.6
2	5.0	4.6	4.1	4.9	4.7	4.7	5.0	4.7	5.0	5.0	4.6	4.8	4.7
3	4.5	4.4	4.0	4.6	4.3	4.2	4.3	4.4	5.0	5.0	4.5	4.3	4.4
4	4.3	4.5	3.8	4.6	4.4	4.4	4.3	4.6	5.0	4.0	4.4	4.5	4.5
5	4.5	4.3	3.3	4.8	4.4	4.5	4.3	4.2	5.0	5.0	3.9	4.4	4.4
6	4.3	4.3	3.7	4.7	4.4	4.5	4.0	3.8	5.0	5.0	3.8	4.2	4.2
7	4.5	4.4	4.2	4.8	4.5	4.3	4.3	4.1	5.0	5.0	4.3	4.3	4.4
8	4.5	4.6	3.0	4.9	4.4	4.3	4.3	4.5	5.0	5.0	4.7	4.5	4.5
9	4.7	4.5	4.7	4.5	4.3	4.5	4.7	4.3	5.0	5.0	4.1	4.5	4.5
10	4.2	4.4	4.7	4.4	4.1	4.1	4.0	4.2	5.0	4.5	3.6	4.4	4.3
11	3.8	4.1	4.5	4.0	3.9	4.1	4.0	4.0	3.0	4.0	3.7	4.0	4.0
12	5.7	4.7	4.3	4.9	4.5	4.9	4.7	4.6	5.0	4.5	4.6	4.5	4.6
Numbers below are percentages													
13a	83	90	83	93	87	75	100	81	100	100	100	86	87
13b	100	89	83	100	94	98	100	94	100	100	100	94	93
14	83	96	100	90	87	80	100	92	100	100	70	84	88
15	17	6	0	33	20	76	33	25	0	100	20	8	39
16a	-	26	17	9	38	23	33	4	-	-	-	30	
16b	100	33	-	40	-	8	-	-	-	-	36	2	
16c	-	41	83	40	62	69	67	96	100	100	64	68	
16d	-	-	-	-	-	-	-	-	-	-	-	0	
17a	-	33	100	17	50	14	67	39	-	50	40	31	35
17b	-	21	-	17	10	14	-	24	-	50	20	16	16
17c	-	-	-	-	10	7	-	-	-	-	-	2	3
17d	100	46	-	66	30	69	33	37	100	-	40	51	46

4. 1998 USAF LABORATORY HSAP MENTOR EVALUATION RESPONSES

Not enough evaluations received (5 total) from Mentors to do useful summary.

5. 1998 HSAP EVALUATION RESPONSES

The summarized results listed below are from the 23 HSAP evaluations received.

HSAP apprentices were asked to rate the following questions on a scale from 1 (below average) to 5 (above average)

1. Your influence on selection of topic/type of work.
2. Working relationship with mentor, other lab scientists.
3. Enhancement of your academic qualifications.
4. Technically challenging work.
5. Lab readiness for you: mentor, task, work plan, equipment.
6. Influence on your career.
7. Increased interest in math/science.
8. Lab research & administrative support.
9. Adequacy of RDL's Apprentice Handbook and administrative materials.
10. Responsiveness of RDL communications.
11. Overall payment procedures.
12. Overall assessment of SRP value to you.
13. Would you apply again next year? Yes (92 %)
14. Will you pursue future studies related to this research? Yes (68 %)
15. Was Tour length satisfactory? Yes (82 %)

	Arnold	Brooks	Edwards	Eglin	Griffiss	Hanscom	Kirtland	Tyndall	WPAFB	Totals
# resp	5	19	7	15	13	2	7	5	40	113
1	2.8	3.3	3.4	3.5	3.4	4.0	3.2	3.6	3.6	3.4
2	4.4	4.6	4.5	4.8	4.6	4.0	4.4	4.0	4.6	4.6
3	4.0	4.2	4.1	4.3	4.5	5.0	4.3	4.6	4.4	4.4
4	3.6	3.9	4.0	4.5	4.2	5.0	4.6	3.8	4.3	4.2
5	4.4	4.1	3.7	4.5	4.1	3.0	3.9	3.6	3.9	4.0
6	3.2	3.6	3.6	4.1	3.8	5.0	3.3	3.8	3.6	3.7
7	2.8	4.1	4.0	3.9	3.9	5.0	3.6	4.0	4.0	3.9
8	3.8	4.1	4.0	4.3	4.0	4.0	4.3	3.8	4.3	4.2
9	4.4	3.6	4.1	4.1	3.5	4.0	3.9	4.0	3.7	3.8
10	4.0	3.8	4.1	3.7	4.1	4.0	3.9	2.4	3.8	3.8
11	4.2	4.2	3.7	3.9	3.8	3.0	3.7	2.6	3.7	3.8
12	4.0	4.5	4.9	4.6	4.6	5.0	4.6	4.2	4.3	4.5
Numbers below are percentages										
13	60%	95%	100%	100%	85%	100%	100%	100%	90%	92%
14	20%	80%	71%	80%	54%	100%	71%	80%	65%	68%
15	100%	70%	71%	100%	100%	50%	86%	60%	80%	82%

The Impact Of Bright Light and a Moderate Caffeine Dose on Nocturnal Performance: A Preliminary Experiment

Harvey Babkoff, Ph.D.
Professor
Department of Psychology
Bar-Ilan University
Ramat-Gan, Israel

Final Report for:
Summer Faculty Research Program
Armstrong Research Site
Associate #98-0094

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Dc

and

Armstrong Research Site

October, 1998

The Impact of Bright Light and A Moderate Dose of Caffeine on Nocturnal Performance: A Preliminary Experiment

Harvey Babkoff
Professor
Department of Psychology
Bar-Ilan University
Ramat-Gan, Israel

Abstract

The impact of the combined and separate effects of a one hour exposure to bright light (~3000 Lux) and to a moderate dose of caffeine on nocturnal performance during a shift-work schedule was studied. Physiological, psychological, performance and biochemical variables were measured during a work-schedule, beginning 1730 in the evening and ending ~1000 the following morning. In the preliminary experiment, five subjects were tested under the four combinations: 1) One hour Bright Light-Placebo; 2) Dim-Light-Placebo; 3) Dim Light-Caffeine; 4) Bright Light-Caffeine. Six other subjects were tested under the "Dim Light-Caffeine" and "Bright Light-Placebo" conditions. This paper reports on the analyses of the impact of the manipulations on the vital signs. Under all experimental conditions, body temperature and heart rate reached their circadian nadir approximately 0430. The reduction in temperature was significantly attenuated by exposure to the bright light. Although the administration of 200 mg caffeine showed a slight trend to attenuate the reduction in body temperature, the change was not significant. For the subjects tested on all four conditions, one hour exposure to an average of 3560 LUX and 200 mg caffeine significantly attenuated the reduction in body temperature relative to the Dim Light-Placebo condition. Changes in heart rate showed a significant circadian rhythm, similar to body temperature and reduced approximately 5 bpm from the evening-night sessions to the late night-early morning sessions. However, in contrast to body temperature, neither exposure to one hour of 3560 LUX nor the ingestion of 200 mg caffeine attenuated the reduction in heart rate associated with the circadian rhythm.

The Impact of Bright Light and a Moderate Caffeine Dose on Nocturnal Performance: A Preliminary Experiment

Harvey Babkoff

Introduction

Bright Lights Several laboratory studies have shown that bright illumination (1000-5000 lux) can improve nocturnal cognitive performance, attenuate the circadian reduction in body temperature and enhance alertness in shift workers and in sleep deprived subjects (French, Hannon & Brainard, 1990; Campbell & Dawson, 1990; Badia Myers, Boecker, Culpepper & Harsh, 1991; French, Whitmore, Hannon, Brainard & Schiflett, 1991; Hannon, Brainard, Childs, Gibson, French, Hanifin & Rollag, 1992). The exposure to bright light is usually prolonged to include all hours of the night or alternations between light exposure and dim light for periods of 2-3 hours each. The effects are reported when subjects are exposed to bright lights continuously during the night, or when the bright lights are alternated with dim lights (Badia et al, 1991). However, some researchers have reported no effect of bright light on temperature when subjects were exposed to bright light throughout a 24-hour period (Daurat, Aguirre, Foret, Gonnet, Keromes and Benoit, 1993). The proposed explanation for this improvement is that the hormone melatonin, normally in abundance during the night, is suppressed by bright light. Since melatonin has been associated with an increase in fatigue (Lieberman, Waldhauser, Garfield, Lynch & Wurtman, 1984), a reduction in melatonin may account for the alerting action of bright light. The suppression of melatonin during exposure to bright light and its recovery after the light is extinguished is related to the intensity of the light (McIntyre, Norman, Burrows and Armstrong, 1989). After exposure to a 3000 LUX light for one hour, melatonin recovered to a 50% level, 35 minutes after the light was extinguished (McIntyre et al, 1989). No studies to date have systematically examined the impact of an exposure of one hour of bright light during a mid-shift break (night shift) on alertness and performance during the subsequent circadian trough (`0300-0600).

Caffeine Caffeine has a general psychostimulant effect in normal conditions as well as after restricted sleep and in irregular work schedules (Rosenthal, Roehrs, Zwyghuizen-Doorenbos, Plath and Roth, 1991). Caffeine reduces reaction time, enhances vigilance performance, increases self-rated alertness and improves mood (Bonnet, Gomez, Wirth and Arand, 1995; Jacobson and Edgley, 1987; Lieberman, Wurtman, Emde and Coviella, 1987; Loke, 1988). The caffeine dose-response curve is non-monotonic. Although caffeine enhances performance at moderate levels (~200-300 mg), there is often a reversal at higher levels (~400-600 mg) (Bonnet et al, 1995; Jacobson and Edgley, 1987). Two hundred mg of caffeine, administered twice during the night, has been reported to suppress melatonin levels and to attenuate the decrease in body temperature (Wright, Badia, Myers, Plenzler and Hakel, 1997).

Modes of Action Bright light is thought to suppress melatonin synthesis by preventing adrenergic stimulation of the pineal gland. There is still some uncertainty about the site and mode of action of caffeine on melatonin. Wright et al (1997) note that adenosine increases melatonin synthesis in the pineal gland. Since caffeine is an adenosine antagonist, its mode of action on mood, performance and body temperature is

thought to be the blockade of adenosine-stimulated melatonin production. Although the site and mode of action of bright lights and caffeine on melatonin may differ, nevertheless, their overall antagonist effect is similar. A recent study (Wright et al, 1997) tested the effects of continuous exposure to a 2000 LUX light and 200 mg caffeine (administered twice during the night), separately and together, on melatonin levels and body temperature in sleep deprived subjects. Either exposure to the bright light or the ingestion of caffeine suppressed melatonin and attenuated the circadian decrease in body temperature. Caffeine together with exposure to the 2000 LUX light suppressed melatonin and attenuated the decrease in body temperature to a greater extent than either condition alone. The effects were additive.

No study to date has examined the combined effects of exposure to bright lights and caffeine on the maintenance of performance levels after the bright light has been extinguished. Nor has any study to date examined the impact of the combination of a moderate dose of caffeine and of an hour's exposure to bright light close to the peak of melatonin synthesis on the maintenance of performance levels during the subsequent circadian nadir (~0300-0600). The present study was a preliminary investigation of the effect of an exposure of one hour of bright light approximately one and a half-hours before the peak of the melatonin secretion curve (around 0300) and 3-4 hours before the circadian nadir on physiological, psychological and performance variables during the circadian nadir. In addition, we examined the combined effects of an hour's exposure to bright light (from 0130-0230) and a moderate dose of caffeine (200 mg) on physiological, psychological and performance variables measured throughout the night, including the circadian nadir.

Method

Bright Light Room We were especially concerned with providing homogeneous exposure of at least 3000 LUX to all of the subjects exposed to the bright light condition for the following reasons: 1) Melatonin suppression is intensity and duration related (McIntyre, Norman, Burrows and Armstrong, 1989); Melatonin recovery begins as soon as subjects leave a bright light environment. Since the experimental design was for only one hour exposure to bright light, there was a definite concern that all subjects receive the maximum available ambient bright light during the hour exposure. To maximize the exposure to light by each subject under the bright light condition, the treatment area (Bright Light Room) was prepared as follows: Six fluorescent fixtures with three bulbs each were hung on the walls at seated-head level. In addition, two large 4-bulb fluorescent fixtures were hung from the ceiling at head level. The walls were painted white and the floor was covered with white linoleum. A white curtain separated the treatment area from the rest of the laboratory testing area. Subjects were seated around a white table. Six subjects were exposed to the Bright Light condition and six subjects were exposed to the Dim Light condition each testing period. Therefore, the ambient luminance at eye level of all of the subjects was measured under a variety of conditions to mimic the possible changes in the light level reaching the eyes when subjects moved their heads. The average luminance levels for six subjects seated around the table ranged from 3050 to 3887.5 LUX, for an average level of 3560 LUX.

Participants Seven men and five women were chosen from among 15 candidates. The participants had no history of chronic illness, sleep disturbances nor drug addiction. The ages of the participants ranged from 19-36 (Av: 24.6 yrs.) and they were mild to

moderate users of caffeine. Eleven of the twelve participants who began the experiment, successfully completed the four weeks of testing.

Experimental Design The "shift-work schedule" consisted of eight testing sessions, the first of which began at 1730, and the last of which began at 0830 the following morning. Session-to-session onset time was two hours, except for the period between the 2330 and 0230 sessions which had the one hour rest period (from 0130-0230) as well.

During the one hour rest period midway in the work-shift, subjects were exposed to one of four experimental conditions: 1) One hour exposure to an average ~3560 LUX from 0130-0230 and 200 mg caffeine at 0140 (Bright Light-Caffeine); 2) One hour exposure to 3560 LUX from 0130-0230 and placebo at 0140 (Bright Light-Placebo); One hour exposure to 20-50 LUX from 0130-0230 and caffeine at 0140 (Dim Light-Caffeine); 4) one hour exposure to 20-50 LUX from 0130-0230 and placebo at 0230 (Dim Light-Placebo). The experimental conditions were separated from each other by one week. During each session the participants performed four different "missions", i.e., computer tests for a total of ~90-95 minutes. Work-to-rest ratio was, therefore, approximately 0.75. All testing took place in dim light (less than 50 LUX). The order of the four experimental conditions was randomized across groups of three participants each over the four test weeks

Because of the preliminary nature of the experiment, six participants were tested under the conditions: Dim Light-Caffeine and Bright Light-Placebo conditions only, while five other participants were tested under all four experimental conditions.

During the week prior to the experiment, subjects participated in orientation and training on the performance batteries which were used in the study.

Testing Procedure Oral temperature, subjective mood and salivary samples were measured approximately once each hour. In addition, salivary samples were also taken once every thirty minutes from 0130 to 0430, for a total of 13 salivary samples per subject per experimental condition. Each testing session included: 1) Two 20-25 minute performance assessment batteries (45 Minute total) measuring working memory load, choice reaction time, spatial discrimination; attention-related performance and logical reasoning; 2) a 28-minute team performance task (DDD); and 3) TRACON, an air-traffic-control task (15-20 minutes). The DDD task requires team interaction (groups of three) and presents a surveillance scenario. At the end of each session participants completed a lighting questionnaire. Beginning 3-days prior to the first experimental condition tested (the first week) and continuing until the end of the last experimental condition tested (the last week), the participants wore actigraphs and complete activity logs.

Data Collection:

Performance Assessment Work Station (PAWS)- The PAWS consists of the following tasks: Grammatical Reasoning (3 min), Matrix Comparison (2 min), Continuous Recognition (2 min), Sternberg Memory Search (2 min), Attention Switching (4 min), Critical Tracking (2 min), Dual Task (2min), and the SAM Fatigue Scale (<1 min).

R&S Performance Battery - Three performance tasks were used. 1) Spatial discrimination at two levels of difficulty. This task presents the subject a target stimulus constructed of 8 adjacent histograms of differing heights. Five comparison stimuli

appear, one of which is different from the target. The subject identifies the different stimulus. 2) Letter Cancellation test working memory at two levels of difficulty. The subject is presented with either 3 or 6 letters on each trial. He must remember the letters and when presented with a matrix of letters, mark the correct letters. 3) Four choice RT tests discrimination using a psychomotor response. Accuracy and speed of responding are measured in all tests.

Dynamic Distributed Decision Making Task (DDD)- This task is a real-time simulation environment in which a complex synthetic team task is implemented. The task will include many of the behaviors that are at the core of the Army participants normal task: assessing a situation, planning response actions, gathering information, sharing information, using resources appropriately in order to accomplish tasks, coordinating actions, and sharing resources. The DDD presents a situation in which there are multiple hostile and friendly synthetic objects. In order to successfully complete their mission, the team of participants will have to gather information about the objects, assess the intentions of the objects, and assign friendly assets to activities which are in accord with the mission (i.e. defend a particular area, investigate a particular 'unknown' object). The participants will work in teams of 3 or 4, depending on the total number of participants available. Various measures will be extracted from this task. Some examples of the measures are: reaction time and accuracy to many different events, resource sharing and team interaction measured in the form of assets shared and overall team performance measured by aggregating individual measures.

Salivary Samples - Participants will be asked to aspirate a 3 cc saliva sample into a syringe which will then be centrifuged and the supernatant frozen. The salivary samples will be analyzed for melatonin content.

Actigraph - The actigraph resembles a wrist watch externally and is worn in a similar manner. Small accelerometers systematically record the subject's movement over time, both while awake and asleep. The more sound asleep, the less a person moves. Thus, actigraphy provides a simple, noninvasive objective means of monitoring the occurrence and depth of sleep.

Activity Log - The activity log was used to provide sleep histories and subjective fatigue ratings for each participant. Participants were asked to indicate their fatigue state every two hours and indicate when they go to sleep and awake.

Oral Thermometer - Oral temperature will be taken during each session. Participants also recorded temperature every two hours during wakefulness for two days prior to and following each session.

Informed Consent: Each participant read and signed a consent form.

Compensation: Participants were paid a total of $500.

Termination: Participants could discontinue participation at any time. One subject chose to terminate his participation after the first week.

Impact of Study

In addition to the theoretical contributions of the research, results of this study may allow military and industrial organizations who employ night shifts to optimally set the lighting levels during the mid-shift break and administer a moderate dose of caffeine for an improvement in nocturnal performance.

Preliminary Results

Body Temperature The average body temperature of the control (Dim Light-Placebo, N=4) and experimental groups (Dim Light-Caffeine and Bright Light-Placebo, N=6) is shown in Fig. 1 as a function of time of day (from 1730-0830). Several points should be noted. 1) The circadian nadir was 0430 for all of the curves. 2) The highest body temperatures at 0230, 0430 and 0630 were recorded for the subjects exposed to the bright light.

The linear component of the slope depicting the decrease in body temperature, from 1930-0430 accounted for approximately 68% of the variance when the temperature data were log-transformed. The slopes of the decrease in body temperature, shown in Table I, indicate that the linear component of the slope of the decrease in body temperature from 1930-0430 for the control group was more than twice as large ($b=-.99405$) as the linear component of the group exposed to bright light ($b=-.4868$) ($F=4.234141$, $p<0.038$). In contrast, the difference in the linear components of the decrease in body temperature between the control group and the group receiving 200 mg caffeine was not significant.

A similar analysis and comparison of the slopes of the decrease in body temperature between 1930-0430 was performed for the five subjects who were exposed to all four experimental conditions. The results are shown in Fig. 2 and in Table II. The shallowest linear slopes were found for the Bright Light-Caffeine condition ($b= -.29405$), which were significantly smaller than those found for the Control (Dim Light-Placebo) condition ($F=6.99042$, $p<0.03829$). None of the other comparisons were significant.

Heart Rate Heart rate displayed a circadian rhythm and decreased approximately 5 bpm from the four evening-night sessions (1730-2330) to the four night-early morning sessions (0230-0830) under all experimental conditions. In contrast to body temperature, there were no significant differences in the slope of the circadian decrease in heart rate across the four experimental conditions.

Preliminary Conclusions

The conclusions must be viewed as very preliminary, given: 1) the small number of subjects tested under all four experimental conditions to date, 2) the fact that the analysis of the saliva-melatonin assays have not yet been performed and 3) that the bulk of the data (psychological and performance) have not yet been analyzed. The following discussion and conclusions are, therefore, limited, and are based only on the results of the vital signs. However, the import of the analysis to date clearly favors the conclusion that the main experimental manipulation was successful. One hour exposure to a homogeneous bright light of ~3560 LUX from 0130-0230 and the administration of a 200 mg dose of caffeine significantly attenuates the reduction of body temperature during the night, even at the circadian nadir. In contrast, the circadian change in heart rate, of approximately 5 bpm, seems to be unaffected by either caffeine, an hour exposure to bright light, or their combination.

Interference in the synthesis of melatonin by a continuous exposure to bright light of 2000 LUX or more has been reported to attenuate the circadian-related decrease in body temperature (Wright, Badia, Myers, Plenzler and Hakel, 1997). Those authors also reported that two administrations of 200 mg caffeine during the night attenuated the reduction of circadian-related temperature reduction. The present findings extend those of Wright et al (1997) and show that even one hour exposure to 3000 LUX during a

mid-night shift rest period was sufficient to attenuate the circadian -related reduction of body temperature for the remainder of the night and early morning. The present findings, thus, emphasize the robustness of the bright light manipulation and the success of the procedure. We must await the melatonin assay findings to determine the percent disruption in melatonin synthesis during the one hour exposure to bright light and the extent to which the exposure to bright light and the administration of 200 mg caffeine continued to interfere with the subsequent recovery of melatonin synthesis after subjects left the bright light room.

The success of the one hour exposure in attenuating the circadian-related reduction in body temperature encourages the further exploration of the combined use of bright light and caffeine in shift-work populations. Clarification of the parameters of bright light, i.e., amount and duration of exposure as well as the required dose of caffeine to effectively attenuate the circadian-related reduction in body temperature is still necessary. Another question regarding the effectiveness of bright light relates to the age of the subjects. The circadian amplitude of Melatonin synthesis is age-related (Tarquini, Cornelissen, Perfetto, Tarquini and Halberg, 1997). Some researchers have reported that the decline in serum melatonin during the early morning hours is more rapid in older than in younger individuals (Ohashi, Okamoto, Uchida, Iyo, Mori and morita, 1997). The results of the present study indicate that even one hour exposure to bright light is effective in attenuating the circadian-related decrease in body temperature in young adults. This raises the question as to whether similar results would be found in older individuals. To date, no systematic study has been performed on the exposure of older individuals to a pulse of bright light during the rise of the melatonin synthesis function and its effectiveness in inhibiting further melatonin synthesis during the night and early morning hours.

References

Badia, P., Myers, B., Boecker, M., Culpepper, J. & Harsh, J. (1991). Bright light effects on body temperature, alertness, EEG and behavior. Journal of Physiological Behavior, 50(3), 583-588.

Bonnet, M.H., Gomez, S., Wirth, O. and Arand, D.L. (1995). The use of caffeine versus prophylactic naps in sustained performance. Sleep, 18, 97-104.

Campbell, S. & Dawson, D. (1990). Enhancement of nighttime alertness and performance with bright ambient light. Journal of Physiological Behavior. 48(2), 317-320.

Daurat, A., Aguirre, A., Foret, J., Gonnet,P., Keromes, A. And Benoit, O. (1993). Bright light affects alertness and performance rhythms during a 24-h constant routine. Physiology and Behavior, 53, 929-936.

French, J., Hannon, P. & Brainard, G. (1990). Effects of bright illuminance on body temperature and human performance. Annual Review of Chronopharmacology Volume 7, France, 37-40.

French, J., Whitmore, J., Hannon, P., Brainard, G. & Schiflett, S. (1991). Photic effects on sustained performance. Fifth Annual Workshop on Space Operations Applications and Research, NASA 3127, 482-486.

Hannon, P., Brainard, G., Childs, R., Gibson, W., French, J., Hanifin, J., & Rollag, M. (1992). Bright light suppresses melatonin and improves cognitive performance during nighttime hours in humans. Proceedings of the 22nd Annual Meeting of the Society for Neuroscience, USA, 195.

Jacobson, B. H. and Edgley, B. M. (1987). Effects of Caffeine on simple reaction time and movement time. Aviation, Space and Environmental Medicine, 58, 1153-1156.

Loke, W.H. (1988). Effects of caffeine on mood and memory. Physiology and Behavior, 44, 367-372.

Lieberman, H.R., Waldhauser, F., Garfield, G., Lynch, H., & Wurtman, R. (1984). Effects of melatonin on human mood and performance. Brain Research. 323, 201-207.

Lieberman, H.R., Wurtman, R.J., Emde, G.G. and Coviella, I.L. (1987). The effects of caffeine and aspirin on mood and performance. Journal of Clinical Psychopharmacology, 7(5), 315-320.

McIntyre, I.M., Norman, T.R., Burrows, G.D. and Armstrong, S.M. (1989). Human melatonin suppression by light is intensity dependent. Journal of Pineal Research, 6, 149-156.

Ohashi, Y., Okamoto, N., Uchida, K., Iyo, M., Mori, N. And Morita, Y. (1997). Differential pattern of the circadian rhythm of serum melatonin in young and elderly healthy subjects. Biological Signals, 6, 301-306.

Rosenthal, L., Roehrs, T., Zwyghuizen-Doorenbos, A., Plath, D. and Roth, T. (1991). Alerting effects of caffeine after normal and restricted sleep. Neuropsychopharmacology, 4(2), 103-108.

Wright, K. P. Jr., Badia, P., Myers, B.L., Plenzler, S.C. and Hakel, M. (1997). Caffeine and light effects on nighttime melatonin and temperature levels in sleep-deprived humans. Brain Research, 747, 78-84.

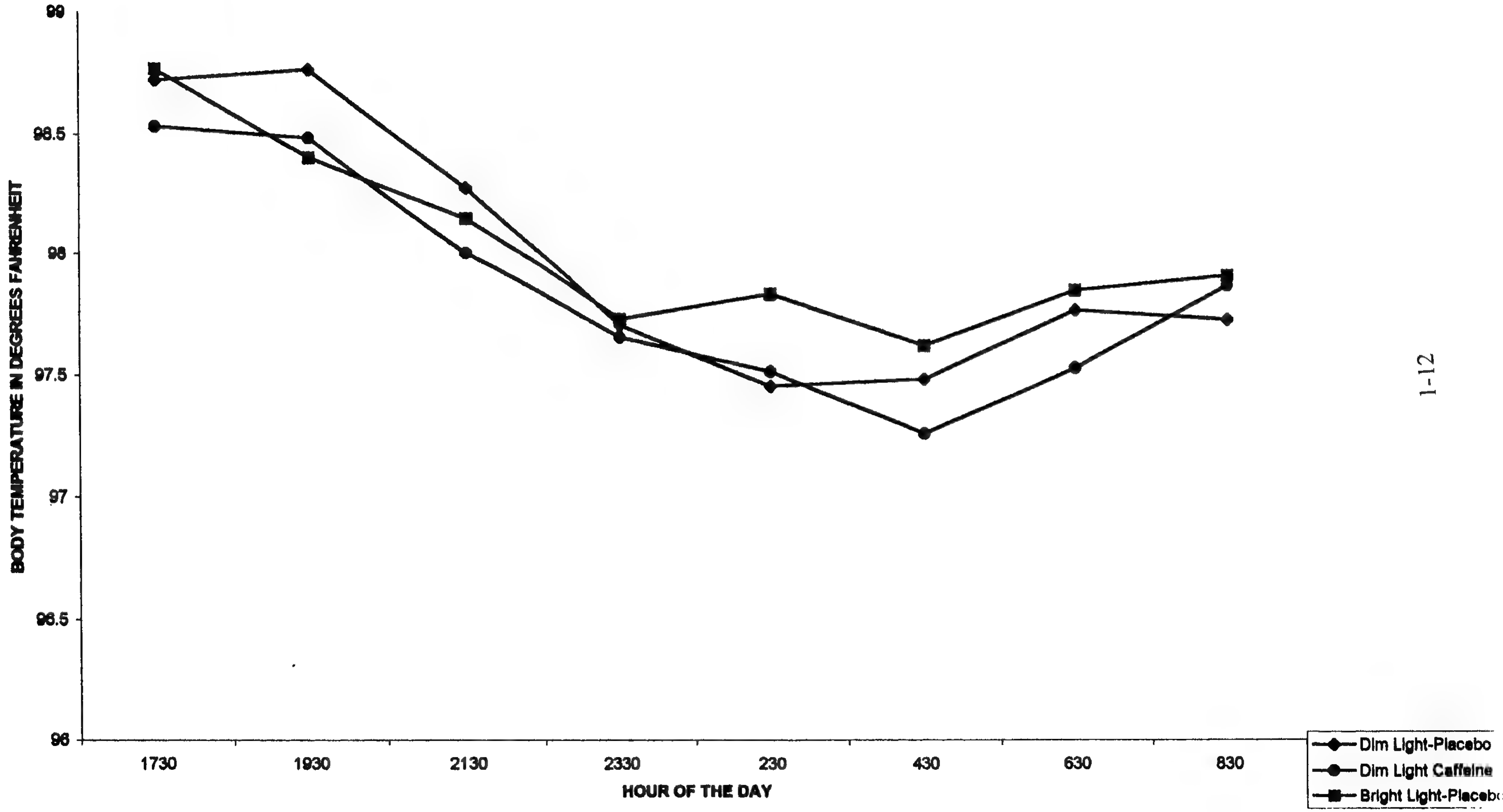
Table I. Average Linear Components of Reduction in Log-Transformed Body Temperature From 1930-0430. Between Groups Comparison

Control Group (Dim Light-Placebo, N=4)	Bright Light-Placebo (N=6)	Dim Light-Caffeine (N=6)
b= -.99405*	b= -.4868*	b= -.74343
95% C.I.= +/-.4402	95% C.I.= +/-.0724	95% C.I.= +/-.08197
* F=4.234141; $p<0.038$		

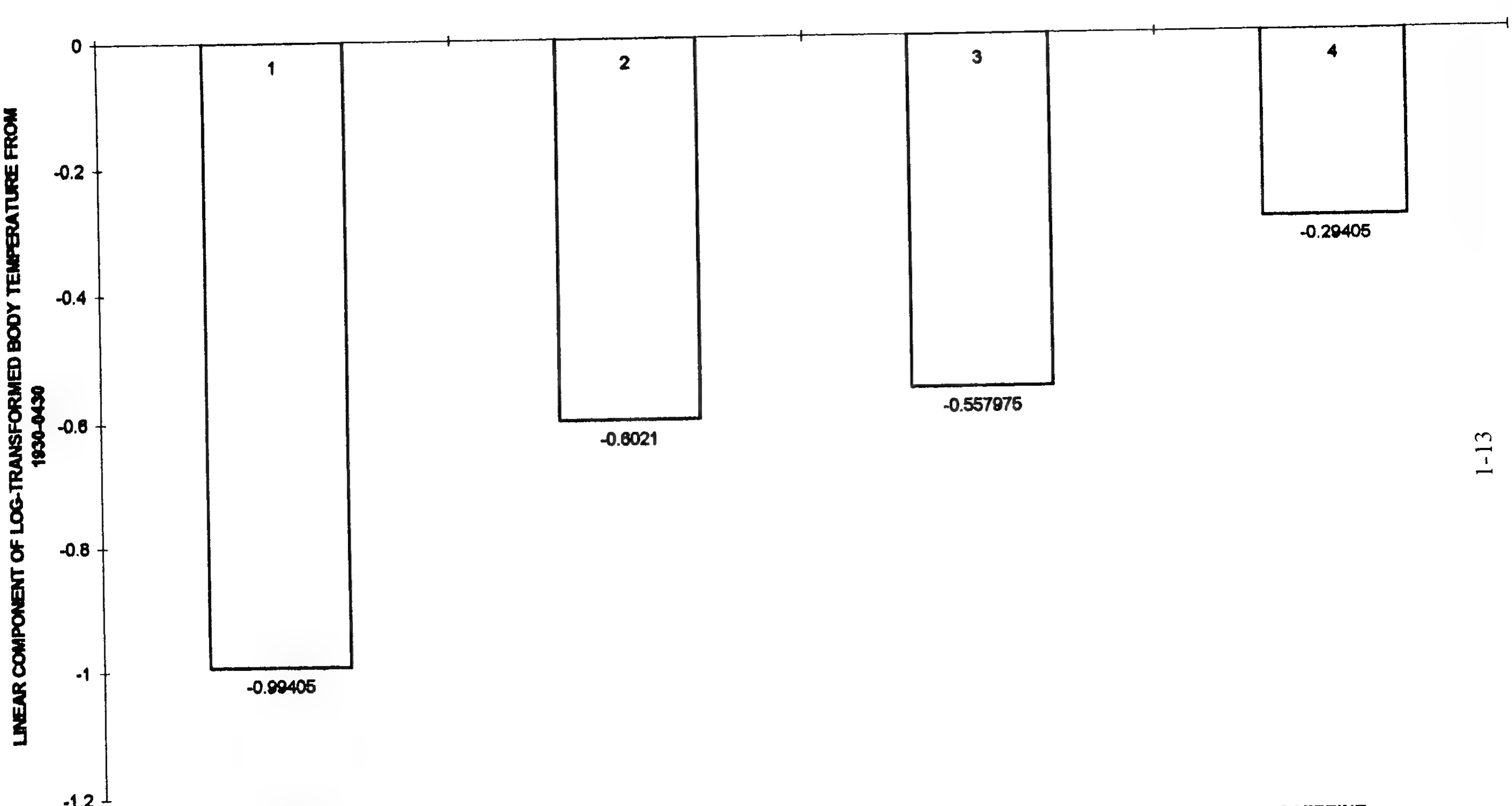
Table II. Average Linear Component of Decrease in Log-Transformed Body Temperature Data from 1930-0430. Within Group Comparison (N=4).

Control Condition: (Dim Light-Placebo)	Bright Light-Placebo	Dim Light-Caffeine	Bright Light-Caffeine
b= -.99405*	b= -.6021	b= -.55798	b= -.29405*
95% C.I.= +/-.4402	95% C.I.= .60858	95% C.I.= .1396	95% C.I.= .1399
*F=6.995042; p<0.038			

BODY TEMPERATURE: COMPARISON OF DIM LIGHT-PLACEBO GROUP WITH DIM LIGHT-CAFFEINE AND BRIGHT LIGHT-PLACEBO GROUP



LINEAR COMPONENT OF DECREASE IN BODY TEMPERATURE. WITHIN GROUP COMPARISON (N=4)



FROM LEFT-RIGHT: CONTROL (DIM LIGHT-PLACEBO);BRIGHT LIGHT-PLACEBO; DIM LIGHT-CAFFEINE; BRIGHT LIGHT-CAFFEINE

1-13

EXPOSURE OF FEMALE RATS TO A 35 GHz ELECTROMAGNETIC FIELD ON DAY 1 OF GESTATION DOES NOT ALTER PREGNANCY RECOGNITION, PREIMPLANTATION EMBRYONIC SURVIVAL, IMPLANTATION RATE, OR EMBRYONIC AND FETAL SURVIVAL TO TERM

Michael P. Dooley, MS, PhD

Affiliate Assistant Professor of Physiology
Department of Biomedical Sciences
College of Veterinary Medicine
Iowa State University
Ames, IA 50011-1250

and

Adjunct Associate Professor
School of Veterinary Sciences
University of Nevada, Reno
Reno, Nevada 89557-0104

Final Report for:
Summer Faculty Research Program

Sponsored by:
Air Force Office of Scientific Research
Bolling AFB, Washington, DC

and

Radiofrequency Radiation Branch, USAF
Directed Energy Bioeffects Laboratory (AFRL/HEDR)
Brooks Air Force Base, San Antonio, TX

November 1998

EXPOSURE OF FEMALE RATS TO A 35 GHz ELECTROMAGNETIC FIELD ON DAY 1 OF GESTATION DOES NOT ALTER PREGNANCY RECOGNITION, PREIMPLANTATION EMBRYONIC SURVIVAL, IMPLANTATION RATE, OR EMBRYONIC AND FETAL SURVIVAL TO TERM

Michael P. Dooley, MS, PhD
Affiliate Assistant Professor of Physiology
College of Veterinary Medicine
Iowa State University

and

Adjunct Associate Professor
School of Veterinary Sciences
University of Nevada, Reno

Abstract

To evaluate whether exposure to a radiofrequency or microwave electromagnetic field (EMF) has the potential to affect developmental processes, studies were initiated using a rat model at the Directed Energy Bioeffects Laboratory at Brooks AFB, TX. Prior studies with this model (Dooley et al., 1996; Dooley et al., 1997) used a microwave transmitter that was capable of generating electromagnetic pulses ranging from near DC to several GHz, but had, as a primary component, microwave pulses in the MHz range. In the present study, a total of 12 male and 26 female rats were obtained from a commercial vendor (Charles River Laboratories, Inc.) and used to determine the effects of exposure of mated females to a 35 GHz electromagnetic field (EMF) on reproductive processes. Females were caged in pairs and estrous females were mated and then allocated to treatment such that females were provided a sham-treatment or exposed to the 35 GHz electromagnetic field (whole body average specific absorption rate = 13 W/Kg) for 25 minutes. Rats were exposed on day 1 of pregnancy. Females were then monitored daily for health and general condition, pregnancy, birth and survival of offspring, and the offspring born to control and treated dams were evaluated for the presence of gross abnormalities and to determine the sex of the feti that were derived from sham-exposed and EMF-exposed embryos. One of 13 rats exposed to the 35 GHz electromagnetic field died during exposure; this death was attributed to the anesthetic protocol. The remaining microwave-exposed rats remained pregnant and produced live offspring at term. Data analysis did not reveal an effect of microwave exposure on the number of females to remain pregnant or the viability and development of embryos, including embryonic survival to term. The proportion of male offspring born to females that were exposed to the 35 GHz electromagnetic field on day 1 or pregnancy was not different from that of the sham-treated females. These results indicate that a 25-minute exposure of recently mated rats to a 35 GHz electromagnetic field at an incident power density of 75 mW/cm^2 was not harmful to the dam or to the establishment and maintenance of pregnancy, when performed under conditions that did not result in hyperthermia of rats (defined here as core temperature < 38 C).

EXPOSURE OF FEMALE RATS TO A 35 GHz ELECTROMAGNETIC FIELD ON DAY 1 OF GESTATION DOES NOT ALTER PREGNANCY RECOGNITION, PREIMPLANTATION EMBRYONIC SURVIVAL, IMPLANTATION RATE, OR EMBRYONIC AND FETAL SURVIVAL TO TERM

Michael P. Dooley, MS, PhD

Introduction

Exposure to electromagnetic radiation is inherent to life within industrialized societies and increasing numbers and types of radiowave and microwave transmitters are being developed and tested with the potential for use in military and/or civilian applications. Some of these systems produce relatively unique forms of electromagnetic field exposure, including electromagnetic pulses at high frequencies, with rapid risetime, and in the case of Ultrawideband (UWB) systems, result in the simultaneous emission of electromagnetic pulses over a broad frequency spectrum.

The effects of microwave radiation on biological systems have been clearly demonstrated for exposures that cause heating of cells and tissues. Unfortunately there has not been consensus regarding cell or animal models to identify and distinguish between thermal and athermal effects or systematic attempts to dissociate the thermal and athermal responses of cells, tissues, and/or animals exposed to microwave sources that cause heating of cells and tissues.

Reproductive function is essential to the continued survival of the species. In recent years, there has been increasing recognition and concern that the decline in the ratio of males to females in developed countries, including the United States and Canada, may be due environmental causes and has led to the proposition that the ratio of male to female births in the population may be a sentinel for harmful effects on the health of man or animals (Davis, **et al.**, 1998). Thus, factors affecting reproductive processes and which have the potential to exert effects on natural biological systems need to be identified. Over the past two decades, the Reproductive Assessment by Continuous Breeding (RACB) model has been used to identify chemical toxicants affecting reproductive processes. Initial studies using the RACB paradigm relied heavily on the mouse as the animal model (Gulati, **et al.**,1991). In recent years, studies have increasingly relied on rats, as the rat model may more correctly identify those human reproductive toxicants that are chemical in nature (Chapin and Sloane, 1997). Unfortunately, testing with the RACB model is expensive to perform and should only be used for the testing of agents when a dose-response relationship can be identified. Hence, this model cannot be justified

and reliably applied for the evaluation of electromagnetic field effects because the responses that have been related to exposures to electromagnetic fields, particularly for athermal exposures, do not evidence a dose-response relationship.

As part of studies performed at Iowa State University, we have proposed (Dooley, **et al.**,1994; Lamont **et al.**,1994) that mammalian embryos at preimplantation stages offer an animal model system in which the cellular effects of exposure to electromagnetic fields can be determined. Further, we have developed a non-toxic assay to complement the morphological evaluation of early embryos (Dooley, **et al.**,1989). The continued viability and development of embryonic cells are essential to the survival of the organism and of the species. Hence, these cells may provide a sensitive and biologically relevant indicator to assess for detrimental consequences of electromagnetic field exposure and to evaluate the safety of new and emerging forms of electromagnetic radiation that are generated as a result of human activity.

This study was performed to obtain preliminary evidence as to whether exposure of female rats to a 35 GHz electromagnetic field on post-mating day 0 affected the development of embryos and feti born to exposed females. Electromagnetic field exposures were applied to the left or right side of pregnant females and in the region overlying the reproductive organs containing the recently fertilized oocytes. Exposures used for this study were shorter than those which are reported to cause circulatory failure and death in rats (Frei, **et al**, 1995) but can be considered to be high-level, radiofrequency (RF) exposures due to their capacity to induce heating in animal tissues (Repacholi, 1998).

Specific objectives of this phase of the study were to: 1) evaluate whether those sham- or 35 GHz-exposed dams, for which the core body temperature did not exceed 38 C during the period of microwave exposure, maintained pregnancy to term and 2) assess whether the controlled application of a microwave field to the skin area overlying the left or right flank exerted local or systemic effects on pregnancy recognition, embryonic survival, or events relation to implantation and fetal survival to term in this animal model.

Materials and Methods

Animals

Sprague-Dawley-derived male and female rats obtained through Charles River Laboratories, Inc., Raleigh, NC were used for this study.

A total of 12 male rats (> 300 g, > 100 days of age) and 26 virgin female rats (> 200 g, > 100 days of age) were available for use in this study. Males were caged individually and provided food and water **ad libitum**. Female rats were caged in pairs and provided food and water as for the males. All rats were housed in plastic cages, kept in the same room, and maintained at 22 C (range 21.8 to 23.2 C) and provided a 14-hour light / 10-hour dark illumination cycle.

Mating and Assignment to Treatment

Female rats were selected for pairing with a male based on vaginal cytology indicative of proestrus or estrus. Female rats were then placed with the designated male rat and left in the male's cage during the dark phase of the illumination cycle. Pairings were sequentially allocated to males such that each would be exposed to a comparable number of sexually receptive females without regard to treatment assignment during the study period. Mating was confirmed on the morning following pairing (day 0) by the presence of spermatozoa in the vaginal smear (Dooley, 1988). Those females for which spermatozoa or spermatozoal remnants could be identified within at least 10 microscopic fields (400X magnification) of the vaginal smear were considered to be pregnant. After mating, females were randomly assigned to treatment (sham- or microwave-exposed) and to the side (left or right) of the animal that was nearest to the millimeter wave source. For each pair of rats, the remaining female in that cage was then allocated to the other treatment group (EMF- or sham-exposed) to ensure that comparable numbers of rats would be assigned to each treatment. Females were allowed to maintain the pregnancy to term to confirm that the embryos within the reproductive tract at the time of microwave-exposure were capable of implantation and survival to term. Data related to the sex-ratio of offsprings born to these sham- and EMF-treated rats were also obtained.

Microwave Source for Exposure of Rats

The 35 GHz exposure source used was a GTE Model 01-1374683-1, Pulsed Millimeter Wave Transmitter. The electromagnetic field created by this Traveling Wave Tube (TWT) amplifier-based transmitter was directed into a temperature-controlled anechoic exposure chamber (Rantec; EMC Test Systems, Austin, TX). The system, capable of an 800 W output, was operated under far-field conditions and at a level such that animals were irradiated at an incident power density of 75 mW/cm^2. The pulse repetition rate was 200 pulses per second (pps) and the pulse width was 14 μsec. The maximum exposure period was 25 minutes for each animal, or until the core body temperature reached 38 C, whichever occurred first.

The output of the microwave source was continuously monitored by the RF technicians assigned to this facility. The millimeter wave was transmitted using a 25 dB standard gain pyramidal horn which was positioned 110 cm from the subject. Free field measurements were performed using a Narda Model 8616 Radiation Survey Meter and a Model 86230 Probe (Narda Microwave Corp., Hanpagga, NY). The power density of the field (forward power of the 35 GHz exposure system) was measured with the aid of a HP 437B Power Meter and a HP R8486A Thermocouple Waveguide Power Sensor (Hewlett-Packard, Corp., Palo Alto, CA).

In conjunction with an infrared (IR) camera, conductive cloth was used to determine the field distribution and the area to position the animals to be treated as part of this study.

Animal Preparation, Microwave Exposure, and Temperature Measurement

To expose rats, animals were anesthetized using Ketamine/Xylazine (70 mg of Ketamine; 10 mg of Xylazine) administered in a single intraperitoneal (i.p.) injection. Females were hair-clipped over the left or right flank and in an area extending from the dorsal to ventral midline. The skin area was then shaved using a razor in the area of the flank and abdomen to be exposed to the electromagnetic field generated by the TWT amplifier.

After preparation of the skin surface, anesthetized rats were transferred from the animal housing facilities to the 35 GHz exposure facility which was located in an adjacent building. Rats were transferred in an enclosed Plexiglas cage, equipped with a lid containing a barrier filter (Allentown Caging & Equipment, Inc., Allentown, NJ) to minimize exposure to potential airborne contaminants. Each EMF- or sham-exposure of an animal was performed within an anechoic chamber. Over the course of this study, chamber temperature was regulated at 24 C and ranged from 22 to 24 C for the exposures provided to the sham-treated rats and from 20 to 25 C for the microwave-exposed rats. Immediately prior to the microwave or sham treatment, the rat was placed on a Plexiglas restraint and positioned on top of a Styrofoam support such that all animals and exposures were performed with animals in the same position relative to the transmission source. Rats were exposed such that either the left or right flank area of the rat was closest to the microwave source (E-orientation). The thermistor probe used to measure core body temperature was taped to the support and to the animal, to ensure that the rat and the probe remained in same position and orientation during the exposure period.

Once each rat was positioned within the exposure facility, the door to the anechoic chamber was closed and animals were provided a 25-minute (maximal) exposure to the 35 GHz millimeter waveform or to a sham-exposure, during which the exposure conditions were identical except no EMF pulses were applied. At the end of the exposure, the female rat was removed from the restrainer, replaced within the transport cage, and then returned to the facility housing the rat colony. Thus, dams were provided a single 25-minute period of exposure to the EMF- or sham-treatment on post-mating day 0, when it was anticipated that embryos at the 1-cell stage would be present within the oviducts.

The whole-body average specific absorption rate (SAR) for rats irradiated using a continuous wave exposure system at an incident power density of 75 mW/cm^2 was determined to be 13 W/kg (Frei, **et al.**, 1995). However, in previous studies with this millimeter wave exposure system, bioeffects related to brief exposures to the 35 GHz field were more closely related to tissue temperature than to the SAR that resulted from exposure of animals to the RF fields. Therefore consideration of heat flow characteristics may be essential, especially for these 35 GHz millimeter wave exposures for which most of the energy is absorbed so near the surface of the animal (NOTE: the estimated depth of tissue penetration of the 35 GHz electromagnetic field is estimated to be <1 millimeter). Thus, significant amounts of electromagnetic energy in the form of heat can dissipate from areas of exposure to high SAR resulting in temperatures within organs and tissues that may be much different from those predicted by schemes that ignore heat flow considerations. For millimeter wave sources, Walters **et al.** (1998) found that the a one-dimensional thermal model based on the solution of the heat conduction equation (Cook, 1952; Foster **et al.**, 1978) compared favorably with measurements of skin temperature as a function of time for short (3-second) exposures using a 94 GHz exposure.

In this study, independent measures of the surface and core body temperature of rats were obtained. Core temperature was measured using a Vitek Model 101 non-metallic thermistor probe inserted through the rectal opening and to extend 4 cm into the colon of the anesthetized rat. Skin temperature was monitored using a Radiance 1 Infrared Camera System (Amber Engineering, Inc., Goleta, CA), equipped with a 50 mm lens. The camera has a focal plane array composed of 256 X 256 indium antimonide sensors and the system was calibrated using a Mikron M340 Black Body Source (Mikron Instrument Company, Inc., Oakland, NJ). Using supporting hardware and software, images were sampled at a rate of 1 per 30 seconds. Off-line image analysis (not included as part of this report) will be performed using ImageDesk™ (Amber Engineering, Inc.) image analysis software.

Experimental Endpoints

The following endpoints were recorded for each pairing of the male with an estrous female rat:

Date of pairing

Presence of spermatozoa in the post-pairing vaginal smear

The following endpoints were recorded for each mated female rat:

Body weight (data not included as part of this report)
- Determined prior to exposure on Day 0
- Determined post-exposure on Days 5, 10, 15, 20, and 25

Duration of exposure to the 35 GHz electromagnetic field (EMF) or to the sham-treatment

Core body temperature (data not included as part of this report)
- Determined at 30 second intervals during millimeter-wave or sham irradiation
- Determined at 30, 60, 90, and 120 seconds after millimeter-wave or sham exposure

Skin temperature at site of irradiation (data not included as part of this report)
- Determined at 30 second intervals during millimeter-wave or sham irradiation
- Determined at 30, 60, 90, and 120 seconds after millimeter-wave or sham exposure

Occurrence and date of parturition
- Number and viability of offspring born at term
- Sex of offspring
- Number of implantation scars evident in the uterine horns at the time of necropsy

Statistical Analyses

Chi-square analysis of ratios was used to compare the proportion of mated females subjected to this millimeter waveform and exposure duration that remained pregnant. Maintenance of pregnancy was based on the identification of implantation scars within the reproductive tracts of the female and/or by the birth of live offspring at term. For those females that produced litters at term, the responses of animals to treatment were compared using Chi-square analysis of ratios for the number of pups born, the number of pups born alive, post-implantation embryonic and fetal losses, and the proportion of male and female offspring. Analyses of variance (ANOVA) were used to compare the litter size and gestation interval for sham-treated, and millimeter wave-exposed rats. Statistical significance was preestablished at $P \leq 0.05$ for all planned comparisons.

Results

The core temperature of female rats at the initiation of the sham or microwave exposure at 75 mW/cm^2 ranged from 33.9 to 36.1 C and for most rats (16/26) was in the range of 34.5 to 35.3 C. The core temperature of sham-treated females continued to decline during the period of sham treatment and by the end of the sham exposure ranged from 31.3 to 33.5 C. For most of the female rats exposed to the 35 GHz electromagnetic field the core temperature declined 0.2 to 0.3 C during the first several minutes of exposure (range 0.2 to 0.8 C) and then heating of the rats by the microwave radiation source caused the body temperature to begin to rise throughout the remainder of the period of microwave exposure. Microwave exposure was discontinued once the core temperature reached 38 C prior to the end of the 25-minute period of exposure (one rat), however for most females the core temperature at the end of the 25-minute period of microwave treatment was in the range of 35 to 36 C (data not shown).

One microwave-exposed female died during exposure. However, the post-mortem examination indicated that the probable cause of death was due to complications related to the anesthesia protocol and was not a consequence of the exposure to the microwave field.

Table 1 summarizes the number of female rats exposed, the side of exposure, and the pregnancy status for sham- and microwave exposures to the left or right flank of the rat. The localized (skin surface) heating caused by the exposure paradigm used for this study did not cause lesions or other signs of skin irritation that could be attributed to the microwave exposure that was administered. Microwave exposure did not interfere with the establishment of pregnancy in these recently mated rats, or affect the survival and implantation of embryos and fetal development to term.

Table 1. **Number of treated females, side of exposure, survival of dams, pregnancy rate, and embryonic and fetal survival to term for sham-exposed dams or for dams that were exposed to a 35 GHz electromagnetic field at 75 mW/cm^2 during day 1 of pregnancy.**

Treatment Group	No. Females	Exposure Applied	No. To Survive (%)	No. Pregnant (%)	No. Litters (%)
Sham-exposed	8	Left	8 (100)	8 (100)	8 (100)
	5	Right	5 (100)	4 (80)	4 (100)
35 GHz-exposed	9	Left	9 (100)	9 (100)	9 (100)
	4	Right	3 (75)	3 (100)	3 (100)

Treatment Group:

No. Females: Number of mated females assigned to treatment.

Exposure Applied: Side of the animal exposed to the 35 GHz electromagnetic field or to the sham-treatment.

No. to Survive: Number of females with a sperm-positive vaginal smear which survived the sham-treatment or exposure to the 35-GHz field for 25 minutes.

No. Pregnant: Number of surviving females for which one or more implantation sites were identified within the uterine horns during post-mortem examination and/or one or more offspring were born at term.

No. Litters: Number of pregnant females for which one or more live offspring were born at term.

Twelve of the 13 mated rats that were exposed to the 35 GHz microwave radiation source survived and of these, 12/12 (100%) maintained pregnancy to term (Table 2). All of the sham-treated rats survived and 12/13 females (92%) maintained pregnancy to term. Necropsy (performed within 200 days of mating) did not reveal any evidence of placental or implantation scars in the uterine horns of the one sham-treated rat which had not produced a litter, suggesting that either ovulatory or fertilization failure or pre-implantation embryonic losses resulted, such that pregnancy could not be established or maintained for this rat.

For the microwave-exposed rats that survived (n = 12), all had implantation sites evident in the near (side that received the direct application of the microwave treatment) and in the far (side contralateral to the microwave treatment) uterine horns. For the sham-exposed rats that maintained pregnancy and were evaluated at necropsy (n = 11); 2/11 (18%) did not present evidence of implantation scars in the near horn (horn ipsilateral to the side of the sham-treatment). Seven of 8 females that were sham-treated on the left side had implantation sites evident in the left uterine horn and 2/3 females that were sham-treated on the right side had implantation sites in the right horn.

All of the rats which revealed evidence of implantation at necropsy had produced live offspring at term (data from Table 1) and the proportion of rats to maintain pregnancy was not affected by the microwave treatment that had been applied to the animal ($P > 0.1$; Table 2).

Table 2. Embryonic survival and maintenance of pregnancy for sham-exposed dams or dams exposed to a 35 GHz electromagnetic field during day 1 of pregnancy.

Treatment Group	Pregnant		Non-Pregnant	
	No. Rats	(%)	No. Rats	(%)
Sham-exposed	12	(92)	1	(8)
35 GHz-exposed	12	(100)	0	(0)

A fertile mating and survival of preimplantation stage embryos was considered to have occurred if one or more implantation sites, based on the presence of implantation scars within the uterine horns, were detected during autopsy.

The proportion of sham-treated female rats to remain pregnant was not different ($P > 0.1$) than for dams exposed to the 35 GHz electromagnetic field on day 1 of pregnancy.

The gestation interval for the sham-treated rats ranged from 21 to 23 days (mean ± SD = 21.9 ± 0.5 days) and from 21 to 22 days for the microwave exposed rats (mean ± SD = 21.8 ± 0.5 days). The proportion of pups born to the sham-exposed rats that survived (born alive) 153/155 (99%) was not different (P > 0.1; Table 3) from that of the EMF-exposed rats for which no dead pups were found.

Table 3. **Post-implantation embryonic and fetal survival to term for sham-exposed dams or dams exposed to a 35 GHz electromagnetic field during day 1 of pregnancy.**

Treatment Group	Born Alive		Born Dead	
	No. Pups	(%)	No. Pups	(%)
Sham-exposed	153	(99)	2	(1)
35 GHz-exposed	168	(100)	0	(0)

The proportion of pups born alive for sham-exposed female rats was not different (P > 0.1) than for pups born to dams that were exposed to the 35 GHz electromagnetic field on day 1 of pregnancy.

The data for the number of embryos to implant, number of offspring born and the estimated number of implanted embryos and/or feti that had died *in utero* and were resorbed prior to parturition is summarized in Table 4. Although twice as many embryos or feti were resorbed in the 35 GHz-exposed rats as compared to the control rats, this difference was not significant (P > 0.1; Table 4).

Table 4. **Number of females, number of embryos to implant, number of offspring identified at term, and estimated incidence of post-implantation losses due to embryonic or fetal resorptions for sham-exposed dams or for dams exposed to a 35 GHz electromagnetic field during day 1 of pregnancy.**

Treatment Group	No. Females*	No. Embryos to Implant	Embryonic / Fetal Development			
			Offspring Born		Embryos / Feti Resorbed	
			No.	(%)	No.	(%)
Sham-exposed	12	147	141	(96)	6	(4)
35 GHz-exposed	12	182	168	(92)	14	(8)

*One 35 GHz-exposed dam died during the period of microwave exposure; one sham-treated female died prior to necropsy and the reproductive tracts were not examined.

No. Embryos to Implant: Total number of implantation (placental) scars evident in the uterine horns of female rats at the time of necropsy.

Offspring Born: Number of offspring delivered at term.

Embryos / Feti Resorbed: Number of implantation sites for which no pup was born alive or dead at term.

The proportion of implanted embryos that were resorbed prior to term for the sham-exposed rats was not different ($P > 0.1$) than for dams exposed to a 35 GHz electromagnetic field on day 1 of pregnancy.

All of the microwave-exposed females that survived the anesthesia and exposure to the microwave field produced a litter. Litter size was not affected by treatment ($P > 0.1$, Table 5) and ranged from 6 to 15 pups for sham-exposures and from 9 to 17 pups for dams exposed to the 35 GHz electromagnetic field.

Table 5. Litter size for sham-treated dams or dams exposed to a 35 GHz electromagnetic field at 75 mW/cm^2 during day 1 of pregnancy.

Treatment Group	Litter Size		
	Mean	(± SD)	Range
Sham-exposed	12.9	(2.9)	6 - 15
35 GHz-exposed	14.0	(2.7)	9 - 17

Data are mean (± SD) for those sham-treated (n = 12) and the 35 GHz-exposed rats (n = 12) that remained pregnant and produced a litter at term.

The mean litter size of sham-exposed female rats was not different ($P > 0.1$) than for females exposed to the 35 GHz electromagnetic field on Day 1 of gestation.

The number of male and female offspring born to rats that were sham- or microwave-exposed on post-mating day 0 (embryos at the 1-cell stage of development) are shown in Table 6. Overall, 83/154 (54%) of the pups born to sham-exposed females were male and the proportion of males that were born was not different ($P > 0.1$, Table 6) from the proportion born to microwave-exposed females (78/167; 47%).

Furthermore, the proportion of litters in which the number of male pups was equal to or exceeded the number of female pups ranged from 8/12 (67%) for the sham-exposure group to 6/12 (50%) for the microwave-exposed females. These differences were also not significant ($P > 0.1$, Table 7).

Table 6. **Number of male and female pups born at term to sham-exposed dams or dams exposed to a 35 GHz electromagnetic field on day 1 of pregnancy.**

Treatment Group	Male		Female	
	No. Pups	(%)	No. Pups	(%)
Sham-exposed	83	(54)	71	(46)
35 GHz-exposed	78	(47)	89	(53)

Data shown are for n = 12 litters for sham-treated and n = 12 litters for the 35 GHz-exposed female rats.

The proportion of male pups born to sham-treated females was not different ($P > 0.1$) than the proportion of male pups born to females exposed to the 35 GHz electromagnetic field on day 1 of gestation.

Table 7. **Proportion of sham-exposed dams or dams exposed to the 35 GHz electromagnetic field during day 1 of pregnancy to produce litters in which the number of males born was equal to or exceeded the number of female offspring for that litter.**

Treatment Group	Male ≥ Female		Female > Male	
	No. Rats	(%)	No. Rats	(%)
Sham-exposed	8	(67)	4	(33)
35 GHz-exposed	6	(50)	6	(50)

The proportion of female rats to remain pregnant and produce litters in which the number of female offspring exceeded the number of male offspring at term was not affected ($P > 0.1$) by the treatment that was provided to the mother on day 1 of gestation.

Three of the 25 rats that survived the experimental exposure protocols used for this study developed skin abscesses in the ventral abdominal region during the period of gestation. This event did not appear to be harmful to the females, had resolved prior to the expected time of parturition, and all of the affected rats had produced litters with live offspring. Two of these rats had been sham-treated and the other was exposed to the 35 GHz field. For these rats, analysis of the data revealed that one sham-treated rat had a litter of 15 pups with no evidence of embryo or fetal resorption prior to parturition and the other sham-exposed female had a litter of 15 pups with 2 resorptions. The microwave-exposed female that developed a skin lesion had 3 resorptions and produced a litter of 9 pups. No evidence of adhesions of the reproductive tissues to the skin or peritoneum or other involvement was noted at necropsy for these or any of the other rats.

Summary and Conclusions

Sources of millimeter waves can result in the exposure of animals and tissues to high frequency electromagnetic fields which are not found in the natural environment (Pakhomov, A. G., **et al.**, 1998). Such waveforms have been shown to induce effects on cells and tissues, and both beneficial and harmful effects in animals and humans have been reported. Microwave exposures have been reported to induce teratogenic effects in rats (Roux, **et al.**, 1986) when core temperatures were significantly elevated (42 C). In mice, a 2.4 GHz field potentiated the effects of a teratogen (cytosine arabinoside), even at nonthermal levels of exposure (Marcickiewicz, **et al.**, 1986). Relatively few studies have been performed in animals using a 35 GHz exposure source and these have focused on the conditions related to lethal heat stress responses of male rats (Frei, **et al.**, 1995; Jauchem, **et al.**, 1997). The reproductive consequences of non-thermal and/or thermal exposures to a 35 GHz microwave field have not been determined in the rat or using other animal models.

Anesthetic agents, such as those used in this study to immobilize rats during the period of microwave exposure, are known to inhibit thermoregulatory responses (Wixson **et al.**, 1987). Hence, in animals and in man, anesthetic-related heat loss has been considered as factor responsible for anesthetic mortality.

For the purposes of the present study, the heat loss that resulted from the anesthesia protocol, when combined with the removal of fur from the skin area selected for microwave irradiation, enabled the microwave exposure of rats to a 35 GHz electromagnetic field at an incident power density of 75 mW/cm^2 for periods of up to 25 minutes, yet under conditions in which the core temperature of the

animals remained hypothermic or thermoneutral. Under these conditions, treatment of rats on day 1 of gestation did not interrupt pregnancy or appear to have an effect on the survival of fertilized oocytes. Furthermore, for these microwave-exposed dams, no other toxicity or harmful effects of EMF exposure were observed in the mother or in the female or male offspring that were born. However, exposures to 35 GHz millimeter waves for longer periods (typically periods of 40 to 50 minutes) and under conditions which result in the substantial elevation of core temperature (~40-41 C) are associated with the death of male rats (Jauchem **et al.**, 1997) and sustained exposures that are known to be lethal can result when skin to core temperature differentials are less than 10 C. Although analysis of the skin temperature measurements for these animals have yet to be performed, no empirical evidence was obtained based on the responses of the animals used in this study to suggest that localized heating of the skin, sufficient to compensate for much of the heat loss caused by the anesthetic agent and to the increased exposure of the skin due to removal of fur, induced localized or systemic effects on reproductive processes related to the maintenance pregnancy or survival of embryos in the rat.

Not all effects of millimeter waves can be explained on the basis of thermal responses, nor do those microwave exposures which result in heating of tissues faithfully mimic thermal responses (Pakhomov, **et al.**, 1998). Hence, exposure conditions and paradigms which do not result in heating of experimental animals (hyperthermia) may provide a useful model for the determination of the effects of prolonged microwave exposures of animals during conditions in which the core temperature remains "thermoneutral"; either due to conditions related to radiant heat loss from the animal or to the disruption of thermoregulatory mechanisms caused by the administration of anesthetic.

In summary, microwave exposure of rats did not affect the viability of embryos, rate of implantation, embryonic and fetal development, and microwave-exposed females produced live offsprings at term. The male and female offspring derived from sham- and microwave-treated rats appeared to have developed normally. Such responses are consistent with the hypothesis that there are no acute or latent effects of exposure to microwave radiation at this intensity, frequency, and for the exposure duration and developmental windows used for these rats, and when the core temperature does not exceed 38.5 C.

The exposures provided to these females were limited to the day of mating and did not continue throughout gestation. Furthermore, exposure were only performed on the embryos resulting in the F_1 generation for these females. Hence the potential effects of exposures to millimeter waves on spermatozoal transport, survival, and/or spermatozoal selection prior to fertilization in mated rats, or to chronic and/or cumulative effects exerted on the oocytes which are destined to be released

in subsequent estrous cycles, as well as other aspects of the reproductive physiology of the male and female are not known.

Acknowledgments

The author gratefully acknowledges the advice and recommendations of Dr. Kathy Ryan, Trinity University and Dr. Thomas Walters, Veridian, Inc., and the technical oversight provided by Mr. William Hurt, AFRL/HEDR. Engineering support was provided by Mr. Leland R. Johnson (Radiofrequency radiation research team leader) and Mr. Luther M. Tate and Mr. Vincent J. Zavala, the Radiofrequency radiation research technicians who administered the 35 GHz microwave exposures of the animals used for this study.

The technical assistance provided by Ms. Terri L. Scholin (Biological effects laboratory team leader) and Mr. Steven J. Dusch (Research Associate, Department of Biology, Trinity University) with the determination of sex-ratio of offspring and to Ms. Jennifer M. Curtis (Biological effects research technician) who participated in the post-mortem evaluation and recovery of tissues from sham- and RF-exposed female rats are also greatly appreciated and hereby acknowledged.

References

Chapin, R. E. and Sloane, R. A, 1997. Reproductive assessment by continuous breeding: Evolving study design and summaries of ninety studies. Environ. Health Perspect. 105 (Suppl. 1):199-205.

Cook, H. F., 1952. A physical investigation of heat production in human tissues when exposed to microwaves. Br. J. Appl. Physics 3:1-6.

Davis, D. L., Gottlieb, M. B., and Stampnitzky, J. R., 1998. Reduced ratio of male to female births in several industrial countries: A sentinel health indicator? JAMA 279:1018-1023.

Dooley, M. P., 1988. The use of eosin B to assess the viability and developmental potential of rat embryos. Dissertation, Iowa State University, Ames, Iowa, 1988.

Dooley, M. P., Pineda, M. H., and Martin, P. A., 1989. Single or intermittent exposure of rat embryos to eosin B does not affect **in vitro** development, survival after transfer, birth of live offspring, and development to weaning. 22nd Ann. Meeting Soc. for the Study of Reprod., Univ. of Missouri, Columbia, MO, August 6-9, 1989, Abstract No. 337. Biol. Reprod. Vol. 40 (Suppl. 1):160.

Dooley, M. P., Pineda, M. H., Lamont, J. W., Weber, R. J., and Moye, D. J., 1994. Continuous Exposure of rat embryos to a 1.5 g electromagnetic field (EMF) does not affect **in vitro** development and viability. Presented at: Experimental Biology 94, Anaheim, CA., April 24-28, 1994, The FASEB Journal Vol. 8, No. 4, Part 1, p. A398, March 1994.

Dooley, M. P., Cobb, B. L., Merritt, J. H., and Jauchem, J. R., 1996. Development of rat embryos exposed to an UltraWideBand (UWB) electromagnetic field. Presented at: Eighteenth Annual Meeting of the Bioelectromagnetics Society, Victoria, British Columbia, Canada, June 9-14, 1996. Session A-8: **In Vivo** Bioeffects, Abstract A-8-1. BEMS Abstract Book, p. 30.

Dooley, M. P., Merritt, J. H., Cobb, B. L., Jauchem, J. R., and Murphy, M. R., 1997. Determination of the influence of ultrawideband exposure of rats during early pregnancy on pregnancy rate, embryonic survival to term, and sex ratio of offspring. Fourth International Symposium on Biologically Closed Electric Circuits, October 26 - 29, 1997, Radisson Hotel South, Bloomington, MN, pp. 268-275.

Foster, K. R., Kritikos, H. N., and Schawn, H. P., 1978. Effects of surface cooling and blood flow on the microwave heating of tissue. IEEE Transactions on Biomedical Engineering 25:313-316.

Frei, M. R., Ryan, K. L., Berger, R. E., and Jauchem, J. R., 1995. Sustained 35-GHz radiofrequency irradiation induces circulatory failure. Shock 4:289-293.

Gulati, D. K., Hope, E., Teague, J. and Chapin, R. E., 1991. Reproductive toxicity assessment by continuous breeding in Sprague-Dawley rats: A comparison of two study designs. Fundam. Appl. Toxicol. 17:270-279.

Jauchem, J. R., Ryan, K. L., Lovelace, J. D., and Frei, M. R., 1997. Effects of esmolol on 35 GHz microwave-induced lethal heat stress. J. Autonomic Pharmacology, 17:165-171.

Lamont, J. W., Weber, R. J., Dooley, M. P., Pineda, M. H., and Moye, D. J., 1994. Shielded culture chamber and controlled uniaxial magnetic field generator for very low frequency (VLF) magnetic field exposure of cells during **in vitro** culture. Presented at: Experimental Biology 94, Anaheim, CA., April 24-28, 1994, The FASEB Journal Vol. 8, No. 4, Part 1, p. A398, March 1994.

Marcickiewicz, J., Chazan, B., Niemiec, T., Sokolska, G., Troszynski, M., Luczak, M., and Szmigielski, .S., 1986. Microwave radiation enhances teratogenic effect of cytosine arabinoside in mice. Biol. Neonate 50:75-82.

Pakhomov, A. G., Akyel, Y., Pakhomova, O. N., Stuck, B. E., and Murphy, M. R., 1998. Current state and implications of research on biological effects of millimeter waves: A review of the literature. Bioelectromagnetics 19:393-413.

Repacholi, M. H., 1998. Low-level exposure to radiofrequency electromagnetic fields: Health effects and research needs. Bioelectromagnetics 19:1-19.

Roux, C., Elefant, E., Gaboriaud, G. Jaullery, C., Gardette, J. , Dupuis, R., and Lambert, D., 1986. Association of microwaves and ionizing radiation: Potentiation of teratogenic effects in the rat. Radiation Research 108:317-326.

Walters, T. J., Blick, D. W., Johnson, L. R., Adair, E. R., and Foster, K. R., 1998. Heating and pain sensation produced in human skin by millimeter waves: Validation of a simple thermal model. (In preparation).

Wixson, S. K., White, W. J., Hughes, H. C., Jr., Lang, C. M., and Marshall, W. K., 1987. The effects of pentobarbital, fentanyl-droperidol, ketamine-xylazine and ketamine-diazepam on core and surface body temperature regulation in adult male rats. Lab. Anim. Science 37:743-749.

KINETIC MODELING OF SLOW DISSOCIATION OF BROMOSULPHOPHTHALEIN FROM ALBUMIN IN PERFUSED RAT LIVER: TOXICOLOGICAL IMPLICATIONS

Brent D. Foy
Assistant Professor
Department of Physics

Wright State University
Dayton, Ohio 45435

Final Report for:
Summer Faculty Research Program
Wright Research Site

Sponsored by:
Air Force Office of Scientific Research
Bolling AFB, Washington DC
and
Wright Research Site

September, 1998

KINETIC MODELING OF SLOW DISSOCIATION OF BROMOSULPHOPHTHALEIN FROM ALBUMIN IN PERFUSED RAT LIVER: TOXICOLOGICAL IMPLICATIONS.

Brent D. Foy
Assistant Professor
Department of Physics
Wright State University

Abstract

Due to strong binding between organic anions and albumin, the kinetics of the binding process must be carefully considered in biologically-based models used for predictive toxicology applications. Specifically, the slow dissociation rate of an organic anion from the protein may lead to reduced availability of free anion in its flow through the capillaries of an organ. In this work, the anion bromosulphophthalein (BSP) was studied in isolated perfused rat livers in the presence of albumin concentrations of 0.25, 1, and 4% (w/v). The uptake of BSP from the perfusion medium was modeled using a biologically based kinetic model of the sinusoidal and intracellular liver compartments. The best fit of the model to data resulted in the prediction of a slow dissociation rate constant for the BSP-albumin of between 0.097 and 0.133 s^{-1}. Assuming BSP and albumin to be in binding equilibrium in the sinusoidal space with rapid binding rate constants, as is often done, produced an unacceptable fit. These results indicate that a strong binding interaction, beyond keeping the concentration of free chemical low due to a small equilibrium dissociation constant, can also reduce uptake by an organ due to the slow release of chemical from the protein during passage through the capillaries. The implication of this dissociation limited condition when extrapolating to other doses and *in vivo* situations is discussed.

KINETIC MODELING OF SLOW DISSOCIATION OF BROMOSULPHOPHTHALEIN FROM ALBUMIN IN PERFUSED RAT LIVER: TOXICOLOGICAL IMPLICATIONS.

Brent D. Foy
Assistant Professor
Department of Physics
Wright State University

Introduction

Many toxins and drugs are transported through the blood with the majority of the chemical bound to a circulating protein. The details of this binding process are important factors in determining the uptake rate and clearance of such chemicals by various organs. When this binding is included in kinetic models, the assumption often made is that the chemical and protein are in binding equilibrium at all times, where the concentrations of free and bound chemical are determined by the equilibrium dissociation constant (K_d). While this assumption will be true for many chemicals, a state of binding equilibrium may not exist for some chemicals (Weisiger et al., 1984; Weisiger, 1985; van der Sluijs et al., 1987; Ott and Weisiger, 1997). A lack of binding equilibrium will cause the concentration of free chemical in the plasma to differ substantially from the concentration of free chemical predicted by using the binding equilibrium assumption. Since the rate of transport of a chemical across a cellular membrane is a function of the concentration of free chemical at the membrane surface for most transport mechanisms, the degree of binding equilibrium can have a major impact on the rate at which a chemical is cleared from plasma and enters cells.

This state of non-equilibrium binding occurs when the rate at which the chemical dissociates from the circulating protein is slower than the rate at which the free chemical is used by subsequent processes such as membrane transport. Thus any free chemical in the plasma is rapidly transported into the cell, but new free chemical in the plasma accumulates slowly due to its slow release from protein. An indication that non-equilibrium binding may be occurring is data demonstrating that the membrane transport or organ uptake rate of a chemical does not correlate well with the predicted (using the binding equilibrium assumption) concentration of

free chemical. Typically, for combinations of binding protein and chemical concentrations in which the equilibrium concentration of free chemical is predicted to be equal, it is found that the uptake rate is dependent on protein concentration and in fact increases for higher concentrations of protein. Chemicals for which this behavior has been seen include bromosulphophthalein (BSP), oleate, and indocyanine green (Weisiger et al., 1984; Ockner et al., 1983; and Ott and Weisiger, 1997).

Previous related studies have focused on the BSP uptake rate for exposure of the elasmobranch liver to a fixed, unvarying concentration of chemical in a single pass perfusion system (Weisiger et al., 1984) or on the 1st pass extraction fraction (Goresky, 1964; Gumucio et al., 1984; Orzes et al., 1985). The study presented here examined the effect of non-equilibrium binding between bovine serum albumin (hereafter referred to as albumin) and BSP on the uptake of BSP by the rat liver, which like other mammalian livers has a much greater BSP uptake rate than the elasmobranch liver. This greater uptake rate has the potential to shift the conditions in which non-equilibrium binding occurs. Also, the experiments were performed in a recirculating perfused liver system with exposure to a single dose of BSP, to simulate a common toxicological kinetic situation. A previously developed biologically based kinetic model for the isolated perfused rat liver (Air Force Technical Report AFRL-HE-WP-TR-1998-0042) was modified to include the association and dissociation rates for chemical-protein binding in the perfusate. The rate of removal of the single dose of BSP from the recirculating perfusion medium for several albumin concentrations was then measured and the data interpreted according to this model. The impact of slow dissociation of chemical from protein on chemical toxicity and safety guidelines established by extrapolation from a limited set of experiments is discussed.

Theory

Protein-binding Kinetics

The kinetic model explicitly includes the rate of association and dissociation for the chemical-protein interaction. This enables the model to simulate situations in which the chemical and protein are not in binding equilibrium in a given compartment, such as the liver sinusoidal

compartment. Assuming a single class of binding sites on the protein, the rate at which the chemical-protein complex is formed (R_{assoc}, units μmole/s) is given by:

$$R_{assoc} = k_{on} \cdot C_{open} \cdot C_{free} \cdot V \tag{1}$$

where k_{on} is the association rate constant ($\mu M^{-1} \cdot s^{-1}$), C_{open} is the concentration of open (unoccupied) binding sites on the protein (μM), C_{free} is the concentration of free chemical (μM), and V is the volume of the compartment in which binding is occurring (L). The rate at which the chemical dissociates from the protein (R_{dissoc}, μmole/s) is given by:

$$R_{dissoc} = k_{off} \cdot C_{bound} \cdot V \tag{2}$$

where k_{off} is the dissociation rate constant (s^{-1}) and C_{bound} is the concentration of the chemical-protein complex (μM). The equilibrium dissociation constant K_d (μM) is then given by:

$$K_d = \frac{k_{off}}{k_{on}}. \tag{3}$$

Liver Uptake

Physiologically, the uptake rate of a chemical from the sinusoidal space into the liver cells is due to the interaction of several processes at the liver membrane and within the cells. The model presented here uses the simplifying assumption that the uptake rate is linearly proportional to the concentration of free chemical in the sinusoidal space. As long as these processes are not close to saturation, then the linear assumption will be valid. Such an assumption has been made before for perfused livers (Weisiger et al., 1984). As explored in the discussion section, the experimental conditions chosen for this study make a linear uptake process likely.

For this linear uptake, the rate at which a chemical is moved from the sinusoidal space to the intracellular space (R_{uptake}, μmoles/s) is given by:

$$R_{uptake} = k_{uptake} \cdot C_{free} \cdot V_s \tag{4}$$

where k_{uptake} is the uptake rate constant (s^{-1}), C_{free} is the concentration of free chemical in the sinusoidal space (μM), and V_s is the volume of the sinusoidal compartment (L).

Kinetic Model

A schematic for the flows and reactions used in the model of the IPRL system is shown in Figure 1. This model was modified from previous work (Air Force Technical Report AFRL-HE-WP-TR-1998-0042) by the addition of the association and dissociation rates of chemical from protein. A brief description of the IPRL model will be presented here.

Perfusion medium is pumped out of the medium reservoir at a constant flow rate. Free chemical, chemical-protein complex, and unbound protein then flow through the sinusoidal space of the liver. The intra-sinusoidal space and the extra-sinusoidal Space of Disse are considered to be a single compartment (called sinusoidal) for both chemical and protein (Goresky, 1980). Within the sinusoidal space, chemical and protein undergo binding reactions at the appropriate association and dissociation rates. Free chemical in the sinusoidal space is then available for uptake into the intracellular space. The subsequent disposition of chemical that is taken into the intracellular space is not modeled. The sinusoidal, intracellular, and reservoir compartments are assumed to be well-mixed. Thus, no concentration gradients exist either along the length of the sinusoid (peri-portal to peri-venous) or transverse to the flow (from cell surface to center of sinusoid).

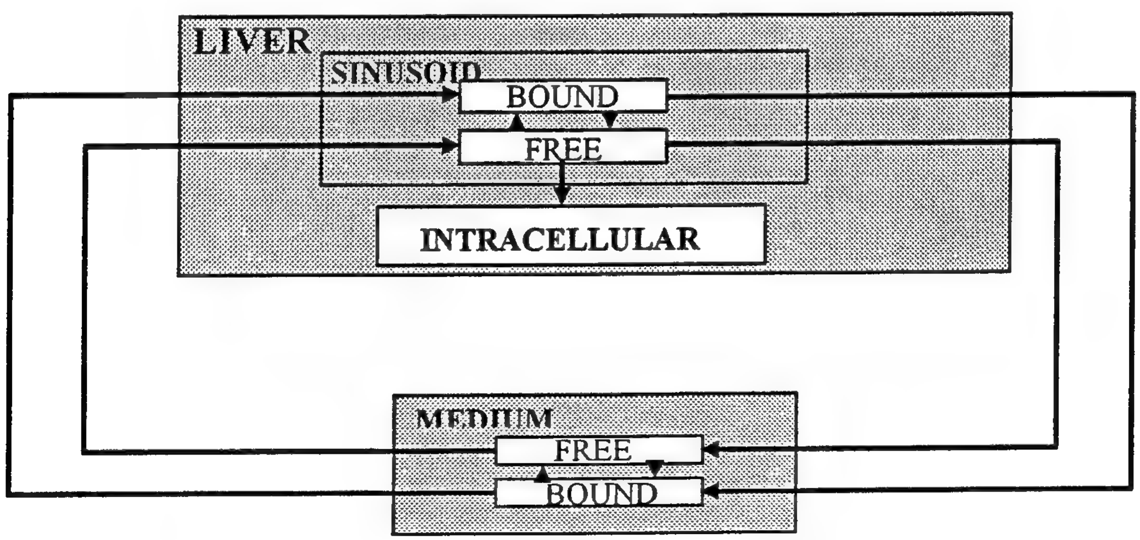


Figure 1. Schematic of the model for the isolated perfused rat liver. The free and bound chemical recirculates through the liver via the medium reservoir. Within the sinusoidal and reservoir compartments, the free chemical can become bound and vice versa. The intracellular space takes chemical only from the free sinusoidal pool.

Methods

Liver Perfusion

Liver surgery and perfusion were performed as previously described (Wyman et al., 1995). Briefly, prior to surgery, male Fisher 344 rats weighing 200-300 g were allowed free access to food and water. After anesthetization with ether and surgically exposing the liver, the bile duct was cannulated, followed by separate cannulations of the portal vein and vena cava. The liver was excised and placed in a dish in a temperature controlled chamber. The flow rate of the perfusion medium was set to 2.4 L/h (40 ml/min) at 37°C, and the perfusion medium was oxygenated by passing it through 25 feet of Silastic tubing exposed to 95% O_2 / 5% CO_2 in a glass chamber. The pH was fixed to 7.4 by adjusting the gas flow rate as necessary. The perfusion medium was Krebs Ringer bicarbonate with 11.5 mM glucose and albumin at 0.25, 1.0, or 4.0% (w/v, low endotoxin, Sigma Chem. Co. cat. no. A2934). The perfusion was conducted in recirculating mode with a total volume of perfusion medium of 200 mL. After 30 minutes for stabilization, 200 μL of 20 mM BSP was added to the reservoir to produce a BSP concentration of 20 μM. A 33.5 mM taurocholate solution (Sigma Chem. Co.) was infused into the perfusion medium at 1 mL/h throughout the experiment to sustain bile production. The total duration of the perfusion after adding BSP was 2.5 h.

At several time points following addition of BSP (2, 5, 10, 20, 30, 60, 90, 120, and 150 min), a 0.5 ml aliquot of perfusion medium was removed from the medium reservoir. The bile outflow was collected over 30 minute time intervals, and bile samples were stored at 0°C until analysis. The concentration of total BSP (BSP plus metabolites such as BSP-glutathione) in perfusion medium was determined by mixing the medium sample with an equal part (0.5 ml) of 1M NaOH and measuring spectral absorbance at 580 nm. For measurement of total BSP in bile, 10 μL of bile sample was added to 1 mL perfusion medium, which was then mixed with 1 mL of 1M NaOH. The spectral absorbance was again measured at 580 nm.

Model Implementation

The biologically based kinetic model was coded on a PC using Advanced Computing Simulation Language (ACSL level 11, MGA Associates, Concord, MA), a numerical integration package.

Model Parameters

The volumes (ml) of the sinusoidal and intracellular compartments are assumed to be 21.7% and 58.3%, respectively, of the weight of the liver (g) (Goresky, 1980).

Regarding binding of BSP to albumin, three studies have found multiple high-affinity binding sites on albumin with K_d less than 0.26 μM. In one study, one binding site had a K_d of 0.06 μM, and two more binding sites with K_d of 0.63 μM (Baker and Bradley, 1966), while a second study found two to three binding sites with a K_d of 0.26 μM (Pfaff et al., 1975). Another study also found a high affinity binding site (K_d = 0.05 μM) but found less than one binding site per albumin molecule at that affinity (Zhao et al., 1993). Since the first two studies both estimate 3 strong binding sites, they are used as a basis for choosing binding parameters. Accordingly, a K_d of 0.2 μM with 3 binding sites per albumin is chosen. Additional lower affinity binding sites on albumin will have a negligible effect on the kinetics due to the low concentration of BSP relative to albumin in all experiments.

Fitting and Error Analysis

These considerations leave two independent parameters to be determined by fitting the reservoir BSP concentration versus perfusion time data: k_{off} and k_{uptake}. The critical information needed to determine these parameters is the relative uptake rates of BSP at different albumin concentrations. Therefore, the data set used for parameter estimation consisted of data from all three albumin concentrations. Application of a non-linear least squares fitting procedure then produced the best estimates of the parameters. In addition, for one fit, k_{off} was fixed to a very large value to explore the kinetics when binding equilibrium is established in the sinusoidal space. Note that setting k_{off} to a large value forces k_{on} to be large due to Eq. 3.

Experimental values are presented as mean ± standard deviation. For fit parameters, 90% confidence intervals were determined by finding the range of each parameter that kept chi-squared within 2.71 of its minimum value (Press et.al., 1992). In other words, the best fit occurs when chi-squared (residuals divided by standard deviations, squared, and summed for each data point) is at its minimum value. The parameter of interest is then fixed to a value that is not its

best fit, and the non-linear least squares algorithm is applied again with all other parameters variable. Chi-squared will be larger, due to the less ideal fit. The 90% confidence interval occurs when a chosen value for the parameter of interest causes chi-squared to increase by 2.71. These confidence intervals are presented in parentheses after the best fit value.

Results

Experimental

Three livers were perfused at each albumin concentration. The average wet liver weights for the perfusions were (in g) 8.63 ± 0.34, 10.39 ± 0.78, and 10.76 ± 1.21, for the 4%, 1%, and 0.25 % albumin concentrations respectively. The volumes of each liver compartment in the model were scaled according to the average weight for each albumin concentration.

BSP concentration in the perfusion medium versus time for each of the 3 albumin concentrations is shown in Figure 2A. In the presence of 4% albumin the concentration of BSP declines the most slowly, which is expected due to the low concentration of free BSP under this condition. The lines in Figs. 2A and 2B are from the modeling discussed below.

Modeling

The results of the modeling are presented as lines in Fig. 2A. The 2 fit parameters are a k_{off} value of 0.114 (0.097 to 0.133) s^{-1}, and a k_{uptake} of 156 (139 to 175) s^{-1}. This k_{off} value, when combined with the equilibrium dissociation constant K_d according to Eq. 3, produces an estimate of the association constant k_{on} of 0.569 (0.486 to 0.667) s^{-1} μM^{-1}. Note that the large value of k_{uptake} relative to k_{off} does not imply that the uptake flux is greater than the dissociation flux since the uptake flux is proportional to k_{uptake} times the extremely low concentrations of free BSP (Eq. 4), while the dissociation flux is proportional to k_{off} times the substantially larger concentration of bound BSP (Eq. 2).

The data was also fit with the assumption that k_{off} equals 2.78 s^{-1} (10,000 hr^{-1}), a large value relative to the 0.114 s^{-1} fit above. This assumption will tend to produce an equilibrium binding situation in the sinusoids. Fig 2B presents the same experimental data as Fig 2A, but the lines represent the predicted reservoir concentration with the equilibrium binding assumption. In this

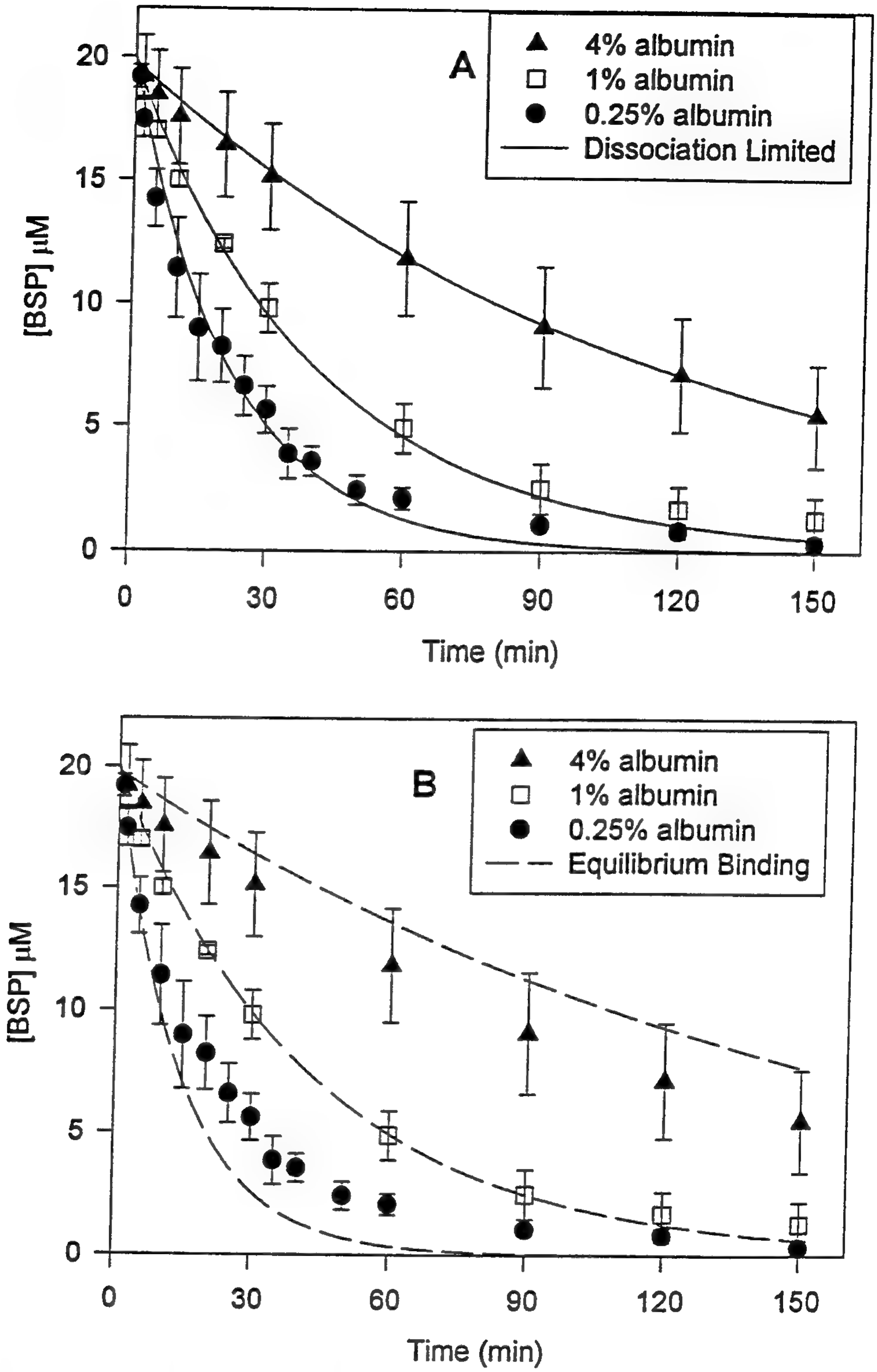


Figure 2. Reservoir concentration of BSP as a function of perfusion time. Symbols represent experimental measurements of BSP concentration, and error bars are standard deviation. **(A)** Dissociation limited case: lines represent best fit prediction from model, with k_{off} estimated to be 0.114 s^{-1} and k_{uptake} estimated to be 156 s^{-1}. **(B)** Equilibrium binding case: symbols represent same data set as in (A). Lines represent prediction from model with k_{off} fixed to 2.78 s^{-1}. k_{uptake} is estimated to be 89 s^{-1}.

case, the data set was fit by varying k_{uptake} alone. The k_{uptake} value for this fit was 89 (81 to 106) s^{-1}. This best fit accurately predicted the 1% albumin data, but was substantially in error for the 4% and 0.25% cases.

To illustrate the prediction of non-equilibrium binding in the sinusoids, Table 1 presents estimates from the model of the free concentration of BSP in the sinusoids when the total BSP in the sinusoid is predicted to be 10 μM. This 10 μM total sinusoidal BSP concentration will occur at a different time for each of the protein concentrations and fitting protocols. For each of the fitting procedures, corresponding to the fits in Figs 2A and 2B, the free concentration of BSP as predicted by the model is compared to the free concentration of BSP that would be present if all the BSP in the sinusoidal space were in equilibrium with albumin. For the dissociation limited fit corresponding to Fig 2A, the predicted free concentration of BSP for the two lower albumin concentrations is substantially different (lower) than the value that would be expected if binding equilibrium existed. Under the fast binding scenario in which k_{off} was set to a high value, BSP is near binding equilibrium for all protein concentrations as expected.

Table 1 – Simulated sinusoidal free BSP concentrations when total BSP in the sinusoid is 10 μM.

	Albumin	0.25%	1%	4%
Dissociation Limited Fit (k_{off} fit to 0.114 s^{-1})	Total BSP (μM)	10.0	10.0	10.0
	Free BSP, Model Prediction (nM)	5.6	2.9	1.0
	Free BSP, Binding Equilibrium (nM)[a]	20.6	4.8	1.2
Fast Binding Fit (k_{off} fixed at 2.78 s^{-1})	Total BSP (μM)	10.0	10.0	10.0
	Free BSP, Model Prediction (nM)	20.3	4.8	1.2
	Free BSP, Binding Equilibrium (nM)[a]	20.6	4.8	1.2

[a]Calculated from albumin and total BSP concentrations assuming 3 binding sites per albumin and a K_d of 0.2 μM.

Another way to distinguish between the purely dissociation limited condition and conditions which may be limited partly by dissociation and partly by intrinsic liver uptake is to compare the relative value of the rate of uptake (R_{uptake}) with the rate of dissociation (R_{dissoc}). From Table 1, the model prediction values for the dissociation limited fit can be used to provide estimates for these rates when total sinusoidal BSP is 10 μM. The values are calculated according to Eqs. (2) and (4), normalized to liver weight for each protein concentration, and presented in Table 2. Note that for 0.25% albumin, R_{uptake} is almost as great as R_{dissoc}. Since the rate at which *free* BSP enters the liver via perfusion is quite small for all albumin concentrations (less than 1% of R_{uptake} per g liver for each data point in Table 2), the majority of free BSP for uptake must come from dissociation. Therefore, R_{uptake} can not exceed R_{dissoc} for any albumin concentration, and the similarity of these values for the 0.25% case indicates that the dissociation rate is the primary factor in determining uptake rate. For 1% albumin, the dissociation rate exceeds the uptake rate by a greater amount, indicating that dissociation is not as limiting in this case. However, the lack of binding equilibrium as shown in Table 1 for the 1% case indicates that dissociation limitations do affect the net uptake rate. Thus the 1% albumin concentration experiment appears to produce an intermediate regime where both the intrinsic uptake rate and the dissociation rate control the net uptake rate. For 4% albumin, the large value of R_{dissoc} relative to R_{uptake} indicates an even smaller importance of dissociation rate in determining net uptake rate. Except for a very small correction due to differing inflow and outflow of free BSP, mass balance considerations force R_{assoc} to equal the difference between R_{dissoc} and R_{uptake}. Thus, in the 4% case, R_{assoc} is close in value to R_{dissoc} since so little of the dissociated chemical is taken up by the cells.

Table 2 – Model predictions for uptake rate and dissociation rate, normalized to average liver weight for each protein concentration. Values in μmoles/(min·g liver).

	0.25%	1%	4%
R_{uptake}/g liver	1.13×10^{-2}	0.58×10^{-2}	0.20×10^{-2}
R_{dissoc}/g liver	1.48×10^{-2}	1.49×10^{-2}	1.50×10^{-2}

Discussion

The experiments and modeling presented here indicate that the rate of dissociation of a toxin from a protein may play an important role in determining the ultimate kinetics for the toxin.

Using the assumption that the dissociation of BSP from albumin was so rapid that equilibrium was maintained in the sinusoids produced an inaccurate prediction of experimental liver uptake kinetics. On the other hand, allowing non-equilibrium BSP binding in the sinusoids enabled the model to match experimental data for 0.25, 1, and 4% albumin while using literature values for the equilibrium dissociation constant.

The ability to extrapolate from a few experimental test cases to a range of possible exposure doses and conditions is a major goal of predictive toxicology. Errors in the underlying extrapolation model can lead to inaccurate predictions. For example, errors in making the assumption of equilibrium binding in the sinusoidal spaces can be seen in Fig. 2B. If the time required for BSP concentration to fall from 20 μM to 3 μM were deemed critical, due possibly to BSP toxicity at a non-liver organ, then the equilibrium-binding prediction of 22 min in the presence of 0.25% albumin is considerably shorter than the experimental finding of 38 min by nearly a factor of two. This may lead to safety exposure limits that are overly risky. The converse problem, predicting clearance by the liver that is slower than reality, will occur if one uses the fast binding parameters at 4% albumin. Although albumin is present in rat plasma at approximately a 4% concentration, under *in vivo* conditions the number of free BSP (or other chemical) binding sites on albumin is likely to be quite variable, due to competing chemicals in the plasma and variable plasma albumin concentrations (under pathologic liver conditions). Thus any predictive model will need to explore a range of chemical:albumin ratios when attempting to model the *in vivo* kinetics of a compound.

Although little data exist regarding the dissociation and association rate constants for the majority of toxic compounds, one can expect that other toxic compounds, especially other anionic compounds, may be so tightly bound to albumin that they behave in a fashion similar to BSP. Using low K_d values as an indication of strong binding and possibly slow dissociation, the phenoxyacetic acid class of pesticides has been shown to bind to bovine serum albumin with K_d values as low as 0.4 μM, which is similar to the K_d for BSP of 0.2 μM (Fang and Lindstrom, 1980). Also, the insecticide chlorpyrifos has been found to bind to albumin with a K_d of 3.4 μM (Sultatos et al, 1984). Plasma carriers other than albumin may also contribute to a potential

dissociation limited condition. For example, alpha 1-acid glycoprotein has been shown to bind the antibiotics lincomycin and clindamycin with K_d values ranging from 1 to 3 μM (Son, et al., 1998).

Some sense of when dissociation limited conditions are likely to affect a toxicokinetic analysis can be gained by identifying such regions on a graph of total concentration of protein binding sites vs. total chemical concentration. At one extreme, dissociation limitations will not occur when the majority of the chemical is free, as opposed to being bound to the protein. When most of the chemical is free, the effect of slow dissociation of the remaining small fraction of bound chemical will be small. The transition zone for the situation in which more than 90% of the chemical is free occurs when:

$$k_{on} \cdot C_{open} < \frac{k_{off}}{9}. \tag{5}$$

At the other extreme, dissociation limited conditions will no longer dominate when binding equilibrium in the sinusoid occurs. Binding equilibrium will occur when the rate of uptake by the liver is smaller than the rates at which chemical associates with or dissociates from the protein. Binding equilibrium will occur at higher concentrations of protein since high concentrations of protein produce low uptake rates (due to low concentrations of free chemical) and high rates of binding. Thus, the transition to binding equilibrium will occur when the rate at which chemical associates with protein is greater than the rate at which chemical is taken up by the liver, or

$$k_{on} \cdot C_{open} > k_{uptake}. \tag{6}$$

After converting the inequalities in Eqs. (5) and (6) to equalities, these boundary lines are graphed in Fig. 3 for a range of albumin binding site concentrations. The rate constants and values for C_{open} used to create the boundary lines in the graph are those determined from the fits and the BSP-albumin binding parameters presented above. These lines are meant to indicate roughly where transitions occur from one behavior to another. When an experiment produces concentrations near one of these boundary lines, the behavior will actually be intermediate in nature.

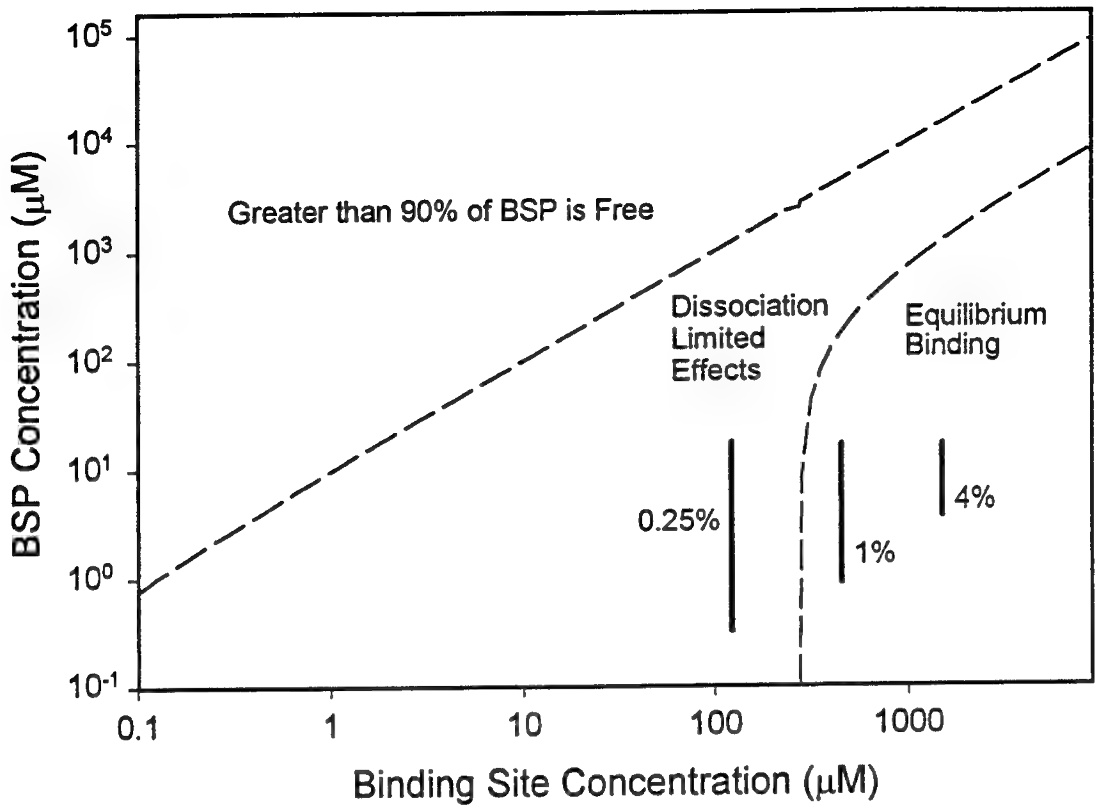


Figure 3. Identification of the chemical and protein concentrations that tend to produce dissociation limited conditions. The dashed lines are mark the boundaries between regions. Solid vertical lines mark the conditions that occurred in experiments in this study. The parameters used to generate these curves are $k_{off} = 0.114\ s^{-1}$, $k_{on} = 0.569\ s^{-1}\cdot\mu M^{-1}$, $k_{uptake} = 156\ s^{-1}$.

One can see that at lower BSP and higher binding site concentrations, the likelihood of dissociation limited conditions is less due to the establishment of near binding equilibrium. Similarly, at high BSP and low binding site concentrations, most BSP will be free and the net uptake of BSP by the liver will be limited by the rate of flow into the sinusoid (perfusion limited). The location of the experiments in this study are also marked on the graph, indicating that the 4% and 0.25% experiments are solidly on either side of the binding equilibrium boundary, while the 1% experiment is near the boundary line (and thus in a transition zone) between dissociation limited and equilibrium binding conditions. The location of the boundary lines will be different for different chemicals, proteins, and organs due to different parameters in

Eqs. (5) and (6). In fact, many chemicals may not exhibit dissociation limited effects for any combination of binding site and chemical concentrations. Dissociation limited effects will be minimal when intrinsic uptake by the liver is slow relative to the rate at which chemical dissociates from the protein binding site. Graphically, this corresponds to a chemical/protein interaction in which the right-hand boundary line (defined by Eq. (6)) has moved to the left, and/or the left-hand boundary line (defined by Eq.(5)) has moved right such that the dissociation limited region no longer exists. For such a chemical/protein interaction, a single transition in the rate-limiting process will occur as the binding site concentration increases, and this transition would be defined not by Eqs. (5) or (6), but rather by the transition from flow-limited conditions to uptake limited conditions (see Weisiger, 1984, for details).

Fig. 4 presents another way to view the dissociation rate limitation on liver uptake. Here, for a given concentration of BSP in the reservoir (10 μM) the rate of uptake into the cells is plotted for a range of concentrations of protein binding sites. Two scenarios are plotted: one which uses the same parameters determined by fitting the experiments in this paper, and one which uses all the same parameters except k_{off} is increased from 0.114 s^{-1} to 2.78 s^{-1} and k_{on} is increased from 0.569 $s^{-1}\cdot\mu M^{-1}$ to 13.9 $s^{-1}\cdot\mu M^{-1}$. These changes keep K_d at 0.2 μM, but setting k_{off} to a high value places the second scenario into a binding equilibrium condition in the sinusoid. The general trend, that uptake rate decreases for increasing binding site concentration, occurs in both scenarios. However, the lower k_{off} in the first case produces a lower uptake rate for a range of binding site concentrations between 1 and 250 μM. The two curves approach each other at low binding site concentrations because both are being limited by the rate at which new chemical is flowing into the liver. At high binding site concentrations, the uptake rate becomes very small due to the small concentration of free chemical, and this low uptake rate enables the chemical and protein to approach binding equilibrium even when k_{off} is small. The data presented in Table 1 also illustrates the phenomenon that higher protein concentrations tend to promote the binding equilibrium condition (Weisiger, 1985; Sorrentino et al., 1994).

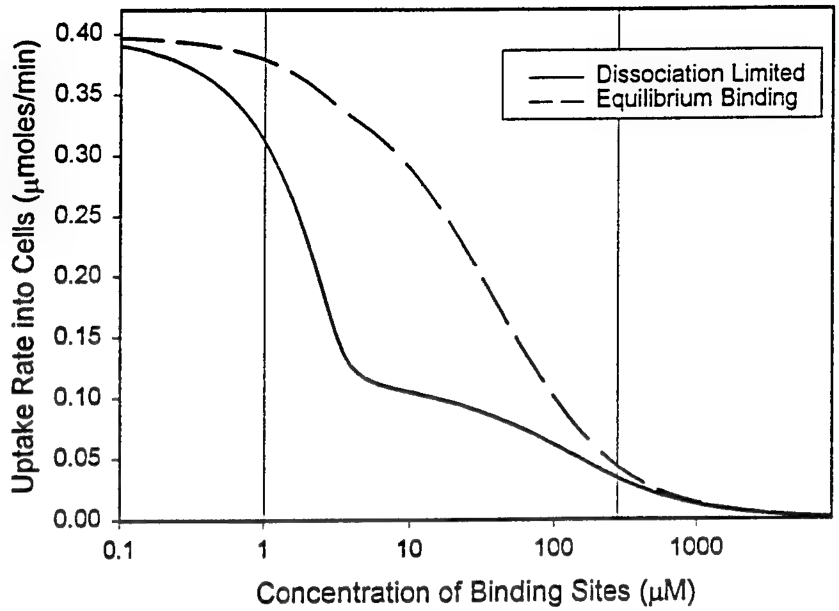


Figure 4. Simulated uptake rate into the intracellular space for a range of binding site concentrations. For these simulations, the initial reservoir BSP concentration was 20 μM and R_i was evaluated at the time points when the reservoir concentration dropped to 10 μM. For the dissociation limited curve, in addition to the parameters listed in Fig. 5, flow Q was 2.4 L/hr and the liver weight was 10g. For the equilibrium binding curve, model parameters were identical except $k_{off} = 2.78\ s^{-1}$ and $k_{on} = 13.9\ s^{-1}\cdot\mu M^{-1}$. The vertical lines correspond to boundary lines identified in Fig. 3 at a BSP concentration of 10 μM.

Regarding the dissociation rate constant determined in this work, no previous direct measurement of k_{off} appears to have been made for BSP-albumin binding. An estimate of k_{off} between 0.053 and 0.208 s^{-1} has been made from application of a model to perfused liver data (Weisiger et al., 1984). The value of 0.114 s^{-1} found in this work is within this range. Measurement of k_{off} for a related compound, dibromosulfophthalein (DBSP), revealed a value of 0.047 s^{-1} (van der Sluijs et al., 1987). Smaller values for k_{off} would tend to promote the dissociation limited condition.

Uptake of BSP by the liver has been treated in a simplified linear manner, with net uptake rate being proportional to the concentration of free BSP in the sinusoidal compartment. The full liver

toxicokinetic model developed in this laboratory (Air Force Technical Report AFRL-HE-WP-TR-1998-0042) takes into consideration several processes not used explicitly in this work, including nonlinear transport at the sinusoidal and biliary membranes and metabolism. Each of these processes has the potential to saturate, producing non-linear effects on the net uptake rate. However, the experimental conditions in this work were chosen to minimize the likelihood of non-linear uptake. By keeping the BSP to albumin ratio low, the amount of free BSP in the sinusoids was kept low, which would tend to keep the membrane transport and subsequent processes operating at rates well below saturation.

The experiments and modeling presented here indicate that a low dissociation rate for a chemical-protein binding interaction can alter the kinetics of the clearance of a chemical by the perfused liver. When using a given set of perfused liver experiments or any limited set of experiments to extrapolate to a wide variety of doses and conditions, ignoring this dissociation limited effect can lead to inaccurate predictions.

Acknowledgments

The authors would like to thank Frank Dessauer and Marcia Feldmann for preparation of isolated perfused rat livers.

References

Baker, K. J., and Bradley, S. E. (1966). Binding of sulfobromophthalein (BSP) sodium by plasma albumin. Its role in hepatic BSP extraction. *J. Clin. Invest.* **45,** 281-297.

Fang, S. C. and Lindstrom, F. T. (1980). In vitro binding of ^{14}C-labeled acidic compounds to serum albumin and their tissue distribution in the rat. *J. Pharmacokin. And Biopharm.* **8,** 583-597.

Goresky, C. A. (1964). Initial distribution and rate of uptake of sulfobromophthalein in the liver. *Am. J. Physiol.* **207,** 13-26.

Goresky, C. A. (1980). Uptake in the liver: The nature of the process. In *Liver and Biliary Tract Physiology*, (N. B. Javitt, Ed.). Vol. 21, pp. 65-101. University Park Press, Baltimore, MD.

Gumucio, D. L., Gumucio, J. J., Wilson, J. A. P., Cutter, C., Krauss, M., Caldwell, R., and Chen, E. (1984). Albumin influences sulfobromophthalein transport by hepatocytes of each acinar zone. *Am. J. Physiol.* **246,** G86-G95.

Koreniewski, C., and Callewaert, D. M. (1983). An enzyme-release assay for natural cytotoxicity. *J. Immun. Methods* **64,** 313-320.

Ockner, R. K., Weisiger, R. A., and Gollan, J. L. (1983). Hepatic uptake of albumin-bound substances: albumin receptor concept. *Am. J. Physiol.* **245,** G13-G18.

Orzes, N., Bellentani, S., Aldini, R., Simon, P., Ferretti, I., Lunazzi, G. C., Sottocasa, G. L., Tiribelli, C. (1985). Sex differences in the hepatic uptake of sulfobromophthalein in the rat. *Clin. Sci.* **69,** 587-593.

Ott, P., and Weisiger, R. A. (1997). Nontraditional effects of protein binding and hematocrit on uptake of indocyanine green by perfused rat liver. *Am. J. Physiol.* **273,** G227-G238.

Pfaff, E, Schwenk, M., Burr, R., Remmer, H. (1975). Molecular aspects of the interaction of bromosulfophthalein with high-affinity binding sites of bovine serum albumin. *Molec. Pharm.* **11,** 144-152.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing.* pp. 656-706. Cambridge University Press, Cambridge.

Son, D. S., Osabe, M., Shimoda, M., Kokue, E. (1998). Contribution of alpha 1-acid glycoprotein to species difference in lincosamides-plasma protein binding kinetics. *J. Vet. Pharmacol. Ther.* **21,** 34-40.

Sorrentino, D., Zifroni, A., Van Ness, K., Berk, P. D. (1994). Unbound ligand drives hepatocyte taurocholate and BSP uptake at physiological albumin concentration. *Am. J. Physiol.* **266,** G425-G432.

Sultatos, L. G., Basker, K. M., Shao, M., Murphy, S. D. (1984). The interaction of the phosphorothioate insecticides chlorphyrifos and parathion and their oxygen analogues with bovine serum albumin. *Mol. Pharmacol.* **26,** 99-104.

van der Sluijs, P., Postema, B., and Meijer, D. K. F. (1987). Lactosylation of albumin reduces uptake rate of dibromosulfophthalein in perfused rat liver and dissociation rate from albumin *in vitro*. *Hepatology* **7,** 688-695.

Weisiger, R. A., Zacks, C. M., Smith, N. D., and Boyer, J. L. (1984). Effect of albumin binding on extraction of sulfobromophthalein by perfused elasmobranch liver: Evidence for dissociation-limited uptake. *Hepatology* **4,** 492-501.

Weisiger, R. A. (1985). Dissociation from albumin: a potentially rate-limiting step in the clearance of substances by the liver. *Proc. Natl. Acad. Sci., USA* **82,** 1563-1567.

Wyman, J., Stokes, J. S., Goehring, M., Buring, M., and Moore, T. (1995). Data collection interface for isolated perfused rat liver: recording oxygen consumption, perfusion pressure, and pH. *Toxicol. Meth.* **5,** 1-14.

Zhao, Y., Snel, C. A. W., Mulder, G. J., Pang, K. S. (1993). Localization of glutathione conjugation activities toward bromosulfophthalein in perfused rat liver. *Drug Metab. Dispos.* **21,** 1070-1078.

NEGOTIATION AT A DISTANCE: WHY YOU MIGHT WANT TO USE THE TELEPHONE

Kenneth A. Graetz
Assistant Professor
Department of Psychology

University of Dayton
300 College Park
Dayton, OH 45469

Final Report for:
Summer Faculty Research Program
Wright Research Site

Sponsored by:

Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Wright Research Site

October 1998

NEGOTIATION AT A DISTANCE: WHY YOU MIGHT WANT TO USE THE TELEPHONE

Kenneth A. Graetz
Assistant Professor
Department of Psychology
University of Dayton

Abstract

Same gender dyads engaged in a 4-issue integrative bargaining task. Negotiator accountability and communication channel (face-to-face, teleconferencing, videoconferencing) were manipulated. Negotiators in the VC condition spent less time negotiating, obtained lower outcomes overall, and engaged in less logrolling that dyads in the other communication conditions. As predicted, negotiators in the TELE condition performed relatively well in the absence of visual access.

NEGOTIATION AT A DISTANCE: WHY YOU MIGHT WANT TO USE THE TELEPHONE

Kenneth A. Graetz

In The Death of Distance, Cairnicross (1997) predicts the emergence of the networked computer as the main platform for supporting global communication and collaboration. This movement toward distributed teams, computer-mediated communication (CMC), and "groupware" is well underway in many organizations and will undoubtedly have profound organizational and psychological consequences. A growing body of theory and research emphasizes the importance of selecting the right tool for the task. For some collaborative activities, certain forms of technological support may be inappropriate and may even introduce new and substantial process losses. For example, a number of studies have demonstrated that synchronous, text-based messaging (i.e., electronic "chat") may be a poor medium for sharing and integrating information in distributed expert groups (Strauss, 1996). Similarly, the expected benefits of electronic brainstorming tools for promoting synergy and enhancing creativity have largely failed to materialize in the laboratory. With the exception of rather large teams (N > 12; Gallupe, Bastianutti, & Cooper, 1991), computer-supported brainstorming groups tend to perform about as well as unsupported groups, who in turn tend to generate as many ideas as isolated individuals, if not fewer (Diehl & Strobe, 1987). Thus, the allure of inexpensive distributed collaboration should be balanced by a firm understanding of the connection between technological support and group task characteristics.

A potential task-technology mismatch involves the use of videoconferencing to support negotiation and bargaining processes. In organizational environments, the growing interest and investment in videoconferencing technology is being driven largely by a desire to "look the other person in the eye." Presumably, this is deemed important when engaging in distributed interactions wherein trustworthiness must be either assessed or communicated. Research suggests that individuals view videoconferencing as more enjoyable, more informative, and more personal than other forms of CMC (Tang & Isaacs, 1993). However, there is also evidence that negotiators use the visual channel to dominate, deceive, threaten, or pressure their opponents and rely heavily on visual cues (e.g., staring) to assess the degree to which an opposing negotiator is trying to dominate them. In a study by Carnevale, Pruitt, and Seilheimer (1981), for example, negotiators who felt accountable to a constituent and who had direct visual access to the opposing negotiator engaged in more contentious behavior and obtained significantly lower outcomes than negotiators who could not see one another. This suggests that the strong preference for videoconferencing as a means of building trust between two distributed parties may be misguided. The current study attempts to replicate and extend the Carnevale, Pruitt, and Seilheimer findings by comparing negotiations in three different communication modalities: face-to-face (FTF), teleconferencing (TELE), and videoconferencing (VC). It was hypothesized that visual access, either FTF or VC, would lead to poor negotiation performance, relatively high levels of frustration and contentious behavior, and low levels of trust. Restricting visual access (i.e., TELE) would presumably reduce the level of contentious behavior and promote interest integration (i.e., logrolling) and the development of trust.

Method

Participants

One hundred and thirty four undergraduates, 68 females and 66 males, participated in the study in partial fulfillment of an introductory psychology research requirement.

Independent Variables

Communication. Dyads were randomly assigned to one of three communication conditions: FTF, TELE, or VC. In the FTF condition, negotiators sat across a table from one another in a private room. Negotiators used notebook computers during the negotiation that did not obstruct negotiators' visual access. In the TELE condition, negotiators were seated in private cubicles (approximately 6 sq. ft), each equipped with a small table, a headset (i.e., hands free) telephone, and a PC. In the VC condition, participants were also seated in private cubicles. The VC connection allowed negotiators to see and hear one another. The VC system provided high quality video and audio over an ISDN line. Video was close to broadcast quality with no noticeable voice lag.

As illustrated in Figure 1, the video window was located in the upper corner of the 17 in color monitor. The cameras were positioned approximately 2.5 ft from the faces of the negotiators and pointed at a slight downward angle. This provided a good view of the opponent's face and upper torso. All discussions were audio taped.

Accountability. Dyads were randomly assigned to either the accountable or not accountable condition. Using a protocol similar to that used in other studies, experimenters told accountable negotiators that their respective "supervisors" would monitor the session. Participants were told that the experimenters would be playing the supervisory roles. Experimenters instructed participants,

> "Although I will not communicate with you at any point during the negotiation, I will be able to hear the discussion. As your supervisor, I will be evaluating your performance as a negotiator using a standard Negotiation Effectiveness Checklist. Based on your performance during the negotiation, you will receive a certain proportion of the total points that result from the final agreement. If I evaluate your performance positively, you will get all of the points. If I evaluate you negatively, you will only get a portion of the points. Remember, as your supervisor I want you to negotiate the best agreement for our side that you can."

In the not accountable condition, the supervisor section of the introduction protocol was omitted. Participants in this condition were simply told to, "Try to negotiate the best agreement for yourself that you can."

Negotiation Task

All dyads engaged in a 4-issue integrative bargaining task similar to those used and described extensively elsewhere (see Pruitt & Lewis, 1975; Thompson & Hastie, 1990; Thompson, 1991). As listed in Table 1, payoff schedules displayed five discrete levels per issue, with points per level listed in parentheses. The task involved the purchase of land from a state land manager by a developer interested in building an amusement park. The negotiation included one compatible issue, the earliest date for opening the park (Opening Day), and one zero-sum issue, the percentage of the park's gross that would go to the state (Gross). Negotiators' point distributions on the compatible issue were identical. For the zero-sum issue, the two negotiators' point distributions, while equal in

magnitude, reflected diametrically opposed preferences. The other two issues, the number of acres sold (Acres) and the percentage of instate park employees (Instate), combined to form a logrolling pair. Although interests on both issues were opposed, each logrolling issue was less important to one negotiator than to the other. The developer valued Acres over Instate, while the buyer preferred Instate to Acres.

As illustrated in Figure 1, all negotiators viewed a computerized version of their payoff schedule, which was located in the lower half of the computer screen. The issue levels and points per level were displayed in menus that dropped down when clicked with the mouse. Negotiators could then click on an issue level and their total points would be calculated and displayed in a box. Thus, negotiators could use this tool to compute the number of points that they would earn for any potential agreement. At no point during the session were the contents of this window transmitted to the other negotiator. Negotiators were only privy to their own payoff schedules.

Procedure

Participants arrived in same-gender pairs. During the recruitment phase, an experimenter ensured that the dyads were unfamiliar with one another. Two experimenters conducted each session. Upon arrival, participants sat at private desks to minimize pre-experimental contact and discussion. After obtaining informed consent, participants read a page of introductory material that described their role, the negotiation issues, and the method of payment. Individuals earned $5 for their participation in the study plus a chance to win a $100 bonus prize. The instructions informed each individual that, following completion of the entire study, the names of individuals who earned a high number of points in the negotiation would be entered into the drawing for the $100. The instructions informed participants that the probability of a tie was quite high and that the recipient of the prize would, "...most likely be selected randomly from among the top point earners." All negotiators were told to, "try to reach a deal that results in the best possible outcome for yourself or for your company."

Experimenters then escorted participants in the FTF condition to the meeting room. Lines of communication were opened between negotiators in the TELE and VC conditions. An experimenter introduced the negotiators to one another by role and advised them as to the 30-min time limit and the consequences of not reaching an agreement. Dyads that failed to reach an agreement within 30 min received 0 points for the negotiation. Experimenters told dyads that a 2-min warning signal would be provided at the 28-min mark. Experimenters informed negotiators that they were not allowed to show their payoff schedules to their opponents during the session.

The experimenters timed the duration of the session. After reaching an agreement, participants in the FTF condition returned to their private rooms. Lines of communication were closed in the TELE and VC conditions. Participants then completed the post-negotiation questionnaire. Finally, experimenters paid and debriefed the participants.

Dependent Variables

Agreements. Analyses were conducted using time to decision and total points across all four issues as dependent variables. In addition, outcomes on the logrolling issues and the compatible issue were analyzed separately. A logrolling score was constructed where 2 = complete logrolling (each negotiator conceded all points on the lesser issue while gaining 4000 points on the more important issue), 1 = partial logrolling (each negotiator

settled for 400 points on the lesser issue while earning 3000 points on the more important issue), 0 = distributive solution (negotiators agreed to split both issues down the middle), –1 = win/lose solution (one negotiator managed to exceed his or her distributive outcome on both issues simultaneously and at his or her opponent's expense), and –2 = lose/lose solution (negotiators reached an agreement that resulted in both obtaining less than the distributive total, essentially logrolling backwards). A score for the dyad was obtained by averaging across negotiators. Performance on the compatible issue was assessed using the total points obtained on that issue.

Post negotiation questionnaire. The post-negotiation questionnaire assessed perceived issue importance and conflict of interest, frustration, trust, and various contentious behaviors. [Note. Specific descriptions of these variables were omitted for the sake of brevity].

Results

Agreements

All of the dyads reached agreement within the time allotted. Averaged across all conditions, the duration of the negotiation was 9.07 min ($SD = 4.47$). A univariate analysis of variance (ANOVA) using time to decision as the dependent variable[1] obtained a significant main effect for communication condition, $F(2, 53) = 4.11$, $p < .05$. Average time to decision (SD in parentheses) was 11.10 min (4.09), 8.98 min (4.77), and 7.13 min (3.72) for the FTF, TELE, and VC conditions, respectively. Pairwise comparisons using a Tukey HSD test revealed a significant difference between the FTF and VC conditions, with negotiators interacting significantly longer in the FTF condition.

Averaging across all conditions, dyads earned 13,400 points[2] ($SD = 1150$), a value significantly greater than the distributive total (12,800), $t(64) = 4.20$, $p < .001$, but significantly less than the integrative total (15,200), $t(64) = -12.61$, $p < .001$. An ANOVA using total points as the dependent variable revealed a significant communication main effect, $F(2, 53) = 3.53$, $p < .05$. The average point total (SD in parentheses) was 13,800 (693), 13,487 (1431), and 12,905 (1035) in the FTF, TELE, and VC conditions, respectively. Pairwise comparisons using a Tukey HSD test obtained a significant difference between the FTF and VC conditions only, with negotiators earning significantly more points overall in the FTF condition. The ANOVA also obtained a significant main effect for gender, $F(1, 53) = 4.62$, $p < .05$. Female dyads earned significantly fewer points overall than male dyads (Ms = 13,006 and 13,697, respectively; SDs = 1191 and 1074, respectively).

Analyzing the logrolling issues separately, 12% of the dyads successfully logrolled, 20% partially logrolled, 17% reached a distributive solution, 45% achieved a win/lose solution, and the remaining 6% arrived at a lose/lose solution. An ANOVA using the logrolling score as a dependent variable obtained a significant main effect for communication, $F(1, 53) = 4.06$, $p < .05$. The average logrolling score (SD in parentheses) was –0.33 (1.06), 0.39 (1.34), and -0.48 (0.93) in the FTF, TELE, and VC conditions, respectively. Pairwise comparisons using a Tukey HSD test obtained a significant difference between the TELE condition and both the FTF and VC conditions. Table 2 lists the percentages of logrolling, distributive, win/lose, and lose/lose solutions by communication condition. Significantly more logrolling occurred in the TELE condition than in either the FTF or VC condition.

Trust

On the post-negotiation questionnaire, participants rated the degree to which they trusted the opposing negotiator (trust-other) and the perceived level of trust afforded them by their opponent (trust-you). The overall means for trust-other and trust-you were 6.79 ($\underline{SD}$ = 1.30) and 6.71 ($\underline{SD}$ = 1.13), respectively. An ANOVA using trust-you ratings as the dependent variable revealed no significant effects. An ANOVA using trust-other ratings as the dependent variable revealed a significant three-way interaction, $\underline{F}(2, 53) = 3.39$, $\underline{p} < .05$, illustrated in Figure 2. Simple comparisons using the pooled error term revealed no significant communication differences for accountable male dyads ($\underline{M}$ = 6.59, averaged across communication conditions). For accountable female dyads, trust-other varied by communication condition, with significantly greater levels of trust reported in the TELE condition versus the FTF condition, $\underline{t}(10) = 3.05$, $\underline{p} < .05$.

Conversely, in the not accountable condition, female negotiators' trust-other levels did not differ significantly across communication conditions ($\underline{M}$ = 7.17, averaged across communication conditions), however, male negotiators' trust-other levels were significantly greater in the TELE condition versus the VC condition, $\underline{t}(8) = 2.61$, $\underline{p} < .05$. The only significant gender difference occurred in the not accountable/VC condition, $\underline{t}(8) = -2.44$, $\underline{p} < .05$, with female negotiators reporting significantly greater trust-other scores than male negotiators.

Contentious Behavior

On the post-negotiation questionnaire, participants rated the extent to which their opponent engaged in a variety of positive and negative behaviors. A series of ANOVAs using each behavior as a dependent variable revealed a number of significant effects. A significant main effect for communication condition emerged for the item "made concessions," $\underline{F}(2, 53) = 3.87$, $\underline{p} < .05$. Means for this item ($\underline{SD}$s in parentheses) were 5.40 (1.30), 6.20 (0.75), and 5.31 (1.40) for FTF, TELE, and VC, respectively. Pairwise comparisons using a Tukey HSD test revealed that negotiators in the TELE condition perceived that their opponent made significantly more concessions than did negotiators in the VC condition.

The ANOVA also revealed significant three-way interactions for the items, "listened to my suggestions/interests," $\underline{F}(2, 53) = 4.42$, $\underline{p} < .05$, and "appeared genuinely friendly," $\underline{F}(2, 53) = 3.33$, $\underline{p} < .05$, with patterns similar to that obtained for trust-other. The means for these items are listed in Table 1.

Correlational Analyses

Table 4 lists the correlations between the post-negotiation items pertaining to opponents' behaviors and both perceived frustration and trust-other. Most of the items were significantly correlated with frustration such that, the greater perceived levels of negative behavior were associate with greater frustration and less trust. Further analyses examining the potential moderational effects revealed that communication condition significantly moderated the relationship between reported frustration and responses on the item, "appeared inflexible," $F(2, 59) = 4.55$, $p < .05$. Although significant correlations were observed between this item and frustration in both the VC, $r(20) = +.68$, $p < .01$, and FTF, $r(20) = +.61$, $p < .01$, conditions, the two were not correlated in the TELE condition, $r(22) = -.08$.

Discussion

Based on the Carnevale, Pruitt, and Seilheimer (1981) findings related to visual access and integrative bargaining, it was predicted that dyads in the TELE condition would have a more accurate impression of their opponent's interests, engage in less contentious behavior, and achieve better joint outcomes than dyads in the FTF

and VC conditions. Although the communication channels investigated in the current study certainly differ along multiple dimensions, the relative absence of major differences between FTF and VC negotiations is a strong indication that the two were qualitatively similar. Thus, it seems unlikely that the poor performance of VC negotiators was due solely to the quality of the visual channel.

Current results indicated that VC may not be the optimal medium for integrative bargaining. Negotiators in the VC condition spent less time negotiating, obtained lower outcomes overall, and engaged in less logrolling that dyads in the other communication conditions. As predicted, negotiators in the TELE condition performed relatively well in the absence of visual access. The study also suggested that the effects of communication condition on perceptions of trust are a complex function of negotiator gender and accountability. Future research should attempt to isolate the mechanisms that may be driving this interaction effect. Finally, the moderational effect of communication condition on the relationship between frustration and perceptions of inflexibility suggests that such perceptions may be strongly influenced derived by visual cues. Thus, one explanation for the superior performance of TELE negotiators may be that they were not exposed to visual indicators of inflexibility and positional commitment.

References

Cairnicross, F. (1997). The Death of Distance. Boston: Harvard Business School Press.

Carnevale, Pruitt, & Seilheimer (1981). Looking and competing: Accountability and visual access in integrative bargaining. Journal of Personality and Social Psychology, 40, 111-120.

Diehl, M., & Strobe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. Journal of Personality and Social Psychology, 53, 497-509.

Gallupe, R.B., Bastianutti, L., & Cooper, W.H. (1991). Unblocking brainstorms. Journal of Applied Psychology, 76, 137-142.

Pruitt, D. G., & Lewis, S. A. (1975). Development of integrative solutions in bilateral negotiation. Journal of Personality and Social Psychology, 31, 621-633.

Straus, S.G. (1996). Getting a clue: The effects of communication media and information distribution on participation and performance in computer-mediated and face-to-face groups. Small Group Research, 27, 115-142.

Tang, J.C., & Isaacs, E. (1993). Why do users like video: Studies of multimedia-supported communication. Computer Supported Cooperative Work, 1, 163-196.

Thompson, L. L. (1991). Information exchange in negotiation. Journal of Experimental Social Psychology, 27, 161-179.

Thompson, L. L., & Hastie, R. (1990). Social perception in negotiation. Organizational Behavior and Human Decision Processes, 47, 98-123.

Footnotes

[1] All statistical tests used $\underline{\alpha} = .05$. For post hoc comparisons, $\underline{\alpha}_{FW} = .05$. All ANOVAs included communication, accountability, and gender as independent variables

[2] Reported point values were rounded to the nearest whole number.

Table 1

Park Developer and Land Manager Payoff Schedules for the 4-Issue Task

Acres	Earliest Opening Day	% of Gross	% Instate Employees
Park Developer's Payoff Schedule			
110 Acres (4000)	1 Year (2400)	2% (2400)	10% (1600)
100 Acres (3000)	2 Years (1800)	4% (1800)	20% (1200)
90 Acres (2000)	3 Years (1200)	6% (1200)	30% (800)
80 Acres (1000)	4 Years (600)	8% (600)	40% (400)
70 Acres (0)	5 Years (0)	10% (0)	50% (0)
Land Manager's Payoff Schedule			
Acres	Earliest Opening Day	% of Gross	% Instate Employees
110 Acres (0)	1 Year (2400)	2% (0)	10% (0)
100 Acres (400)	2 Years (1800)	4% (600)	20% (1000)
90 Acres (800)	3 Years (1200)	6% (1200)	30% (2000)
80 Acres (1000)	4 Years (600)	8% (1800)	40% (3000)
70 Acres (1600)	5 Years (0)	10% (2400)	50% (4000)

Note. Points for each issue are listed in parentheses. All payoff schedules used in the study presented levels for both players arranged from most to least preferred.

Table 2

Logrolling Performance by Communication Condition

Outcomes	FTF	Telephone	Videoconference
Complete Logrolling	9%	26%	0%
Partial Logrolling	14%	26%	19%
Distributive Solution	9%	17%	24%
Win/Lose Solution	68%	22%	48%
Lose/Lose Solution	0%	9%	9%

Table 3

Mean Responses to Contentious Behavior Items by Communication, Accountability, and Gender

	Not Accountable		Accountable	
	Male	Female	Male	Female
FTF				
Trust-Other	6.25	7.50	6.60	6.42
Listened-Other	6.58	7.88	7.70	6.83
Friendly-Other	6.92	7.50	7.40	7.08
Friendly-Own	7.33	7.25	7.60	7.33
Telephone				
Trust-Other	7.58	6.70	6.67	7.67
Listened-Other	8.08	7.30	7.75	8.08
Friendly-Other	7.83	6.80	7.08	8.42
Friendly-Own	7.75	7.30	6.75	8.42
Videoconference				
Trust-Other	5.75	7.33	6.50	6.33
Listened-Other	6.97	7.27	8.20	7.67
Friendly-Other	6.38	7.75	8.00	7.67
Friendly-Own	5.83	7.92	7.60	7.25

Table 4

Simple Correlations of Behavioral Items With Frustration and Trust-Other

Items	r(frus)	r(trust)
Made concessions	-.32*	+.29*
Communicated interests	-.14	+.41*
Offered Alternatives	-.43*	+.40*
Interested in reaching agreement	-.50*	+.60*
Listened to my suggestions	-.26*	+.50*
Appeared friendly	-.35*	+.66*
Purposely misled	+.39*	-.50*
Used verbal pressure tactics	+.29*	-.28*
Used nonverbal pressure tactics	+.42*	-.50*
Appeared inflexible	+.40*	-.46*
Appeared mainly interested in self	+.23	-.27*

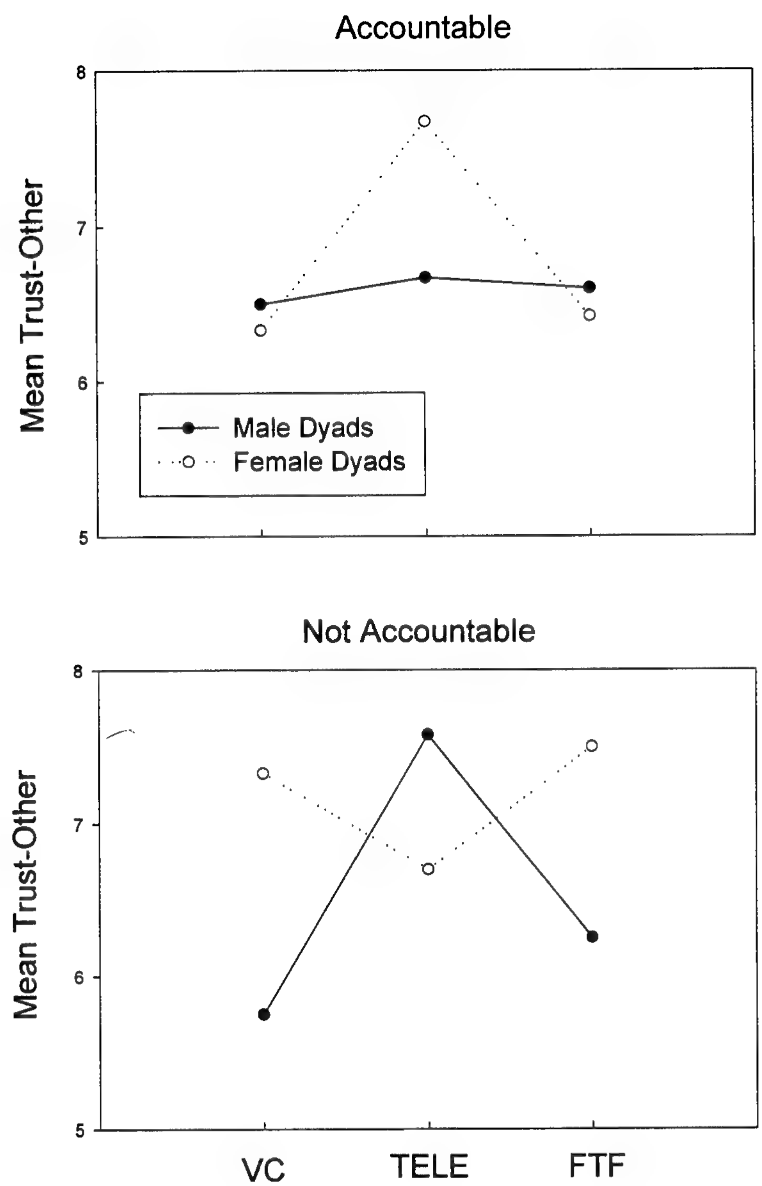


Figure 1. Mean Responses on Trust-Other by Communication, Accountability and Gender

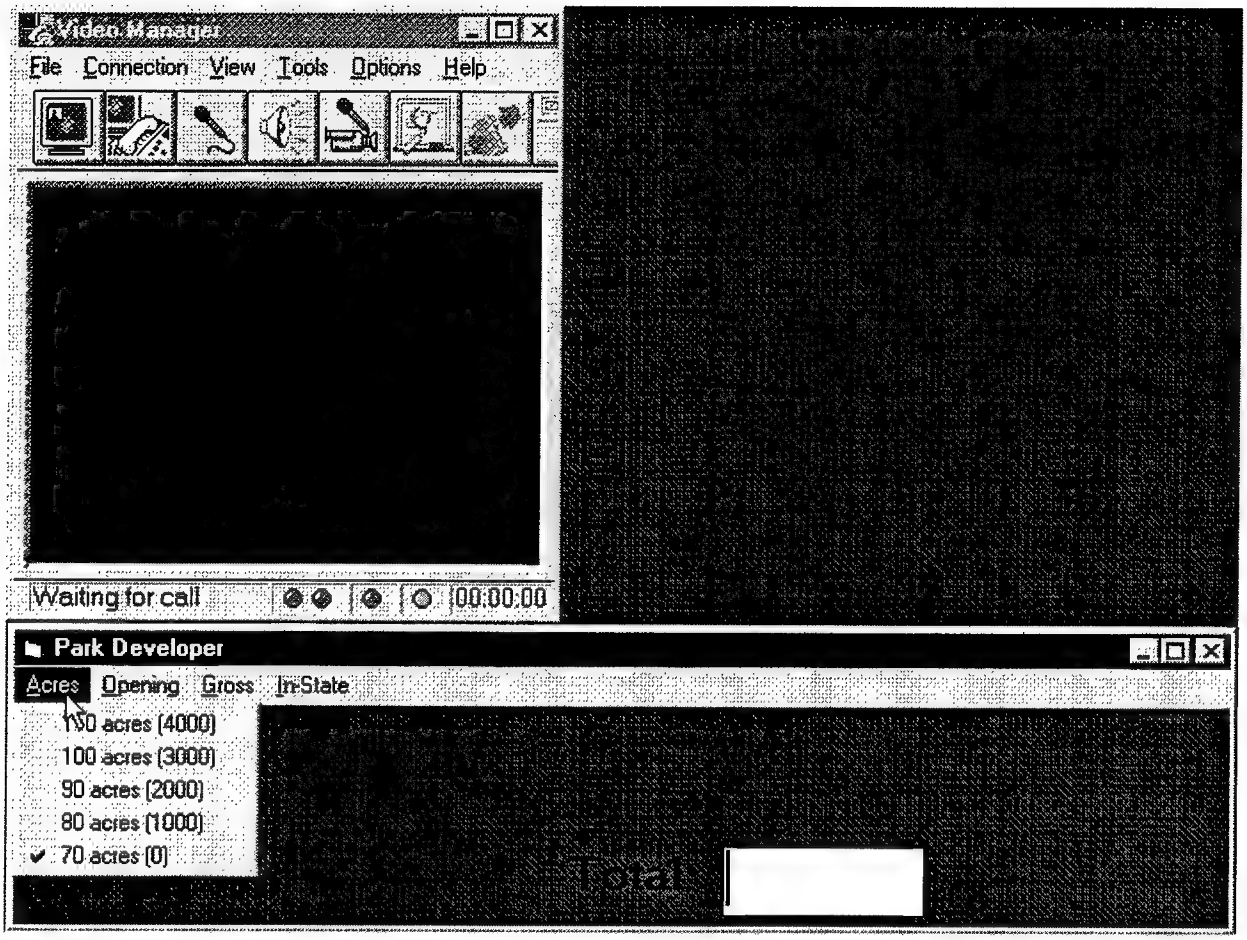


Figure 2. Screen Shot of Video Window and Computerized Payoff Schedule

Conceptualizing Crew Performance in

Dynamic Operational Environments:

A Hierarchy of Embedded Action-Control Models

Verlin B. Hinsz, Ph.D.
Professor of Psychology
Minard Hall 115
North Dakota State University
Fargo, ND 58105

Final Report for the Summer Faculty Research Program

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Crew System Interface Division
Human Effectiveness Directorate
Air Force Research Laboratory
Wright-Patterson Air Force Base, Ohio

August 1998

ABSTRACT

A conceptual framework is proposed for modeling crew performance in dynamic operational environments. This framework evolves from models of individual performance on dynamic tasks that focus on goal-directed behavior. The conceptual framework also builds upon theoretical developments in control-systems theory and groups as dynamic systems. A series of models are developed that represent (1) the processes of crew members performing their tasks, (2) the processes of interaction among the crew members, and (3) the processes of the crew as a functioning unit. These models form a hierarchy of embedded systems for the dynamic processes in the operational environment. The central concepts for crew performance in these operational environments are the interdependence among the members, and the strategies for coordination and collaboration. An illustration of the use of these models of crew performance is made based on activities of crew members for a Uninhabited Aerial Vehicle ground control station. Subsequently, the models of crew performance are used to highlight a number of research agendas that can contribute to our understanding of crew performance in dynamic operational environments such as Uninhabited Aerial Vehicles.

Conceptualizing Crew Performance in Dynamic Operational Environments:

A Hierarchy of Embedded Action-Control Models

As we approach the beginning of a new century, the Air Force is exploring new ways of fulfilling its mission in defense of our nation. One perspective on how the Air Force can be more effective is to develop strategies and technologies that enhance the way that Air Force crews collaborate as they perform the complex tasks their members face. Many of these crews perform these tasks in rapidly changing environments. Inherent in crew performance of dynamic, complex tasks is the concept of interdependence. As the interdependence of crew members increases (in terms of goals, actions, outcomes, interactions, and rewards), effective strategies of coordination and collaboration become more important for superior performance. This paper develops a conceptual framework for considering crew performance in a complex and dynamic operational environment (i.e., reconnaissance UAV). This conceptual framework builds upon theoretical developments in information processing in teams, control-systems theory, and groups as dynamic systems. The conceptual framework then leads to an outline of a number of research agendas regarding performance in UAV crews.

Crew Performance in a Dynamic Operational Environment

The conceptual framework developed here is based upon a number of theoretical and conceptual bases. Because the ultimate objective is to develop a model of crew performance in a dynamic operational environment, this model will rely substantially upon models of individual action in dynamic operational environments. In particular, a dynamic model of performance for goal-directed behavior appears appropriate. However, for considerations of crew performance in dynamic environments, it is necessary to go beyond the actions of individuals to consider the nature of interdependence and how it leads to coordination and collaboration for effective crew performance. Coordination and collaboration become more important for effective performance as tasks become more interdependent, complex, and dynamic, and as the interactions among crew members are longer in duration, have greater intensity, and occur more frequently.

The conceptual framework developed in this paper also rests upon recent theoretical developments that consider groups as complex, adaptive, and dynamic systems (McGrath, Arrow, & Berdahl, 1999). These theoretical developments go beyond traditional research on

crews' one-time performance in operational environments to situations in which the crew members adapt to changes in the task environment and changes in the other crew members. Crews performing these tasks expect and anticipate changes in the task, and expect to continue interaction on tasks in the future. Moreover, crews can be seen as dynamic collectives that respond cognitively and socially to the rapidly changing aspects of the situations they face. The conceptual framework presented in this paper goes beyond these important recent theoretical developments by focusing on a hierarchy of embedded systems. This aspect of the conceptual framework is an elaboration of control and systems models of individual performance on dynamic tasks. Consequently, the development of the conceptual framework will begin with a description of a useful model of individual task performance on dynamic tasks.

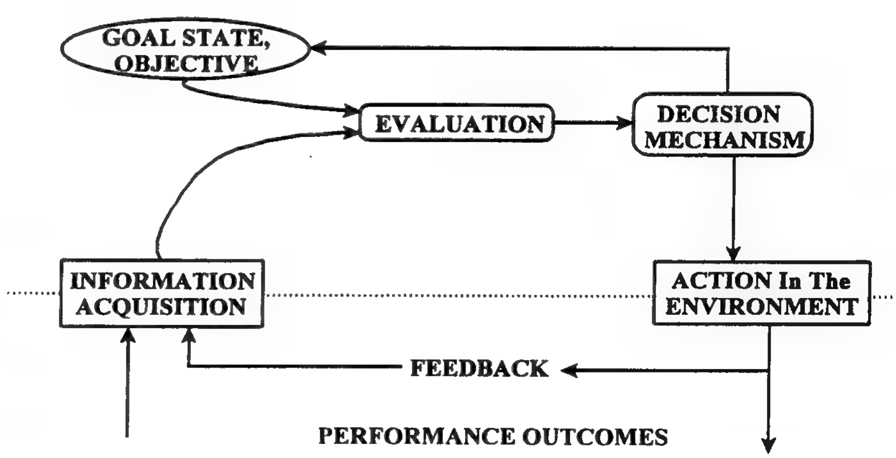
A Model of Individual Performance in Dynamic Operational Environments

There is a well-developed literature regarding the performance of individuals on dynamic tasks (e.g., Smith, 1996). Consequently, a number of models of individual performance are available that can serve as foundations upon which to build a crew-based model. One constraint on the type of model considered is that it should focus on goal-directed behavior leading to performance. The performance of individuals on dynamic tasks of interest would involve the pursuit of some objectives or mission.

A general approach to performance in dynamic environments that focuses on goal-directed behavior is that of control theory (i.e., Carver & Shreier, 1981; H. Klein, 1989; Lord & Kernan, 1989; Powers, 1973; Rasmussen, Pejtersen, & Goodstein, 1994). Figure 1 presents

Figure 1

A simple feedback-loop model for actions performed on a task.



a general version of a simple feedback-loop model of goal-directed behavior related to task performance which is an extension of control-systems theoretical approaches. The representation in Figure 1 closely resembles that of Lord and Kernan (1989) in their model of motivation for goal-directed behavior for task performance. This model will serve as a basis upon which to begin a consideration of performance on dynamic tasks.

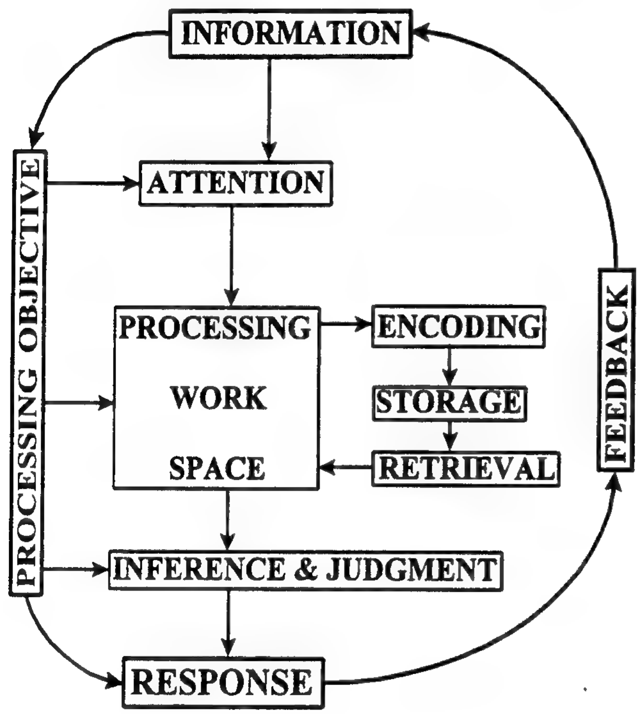
This simple feedback-loop model divides the realm of interest into the portion that is in the human task performer (above the dotted line) and a portion in the environment in which the task exists (below the dotted line). In this system, the human acquires relevant information from the environment. This information is usually relevant to some objective or desired goal state. A comparison between the state of the environment as perceived and the desired goal state occurs, and an evaluation of any discrepancies is made. If the consequence of the evaluation is that some discrepancy exists requiring action, the decision mechanism then decides what type of action is required. The "action" can be of various kinds. The decision mechanism may choose a behavioral change regarding the nature of the action (intensity, direction, persistence). The decision mechanism may also lead to cognitive changes (e.g., a change in the goal, a change in a frame, redirect the search for information). These decisions are then translated into actions on the part of the person in the environment. These actions and other events will then result in changes in the environment, which will serve as new information that may be acquired.

The actions the individual produces are generally directed toward the task of interest. These actions should influence the task and contribute to the critical performance outcomes that are assessed. The model presented in Figure 1 implies that the action has this impact on the task resulting in the performance outcomes that can be measured. The relation between action and performance outcomes is complex and indirect (Kanfer, 1992). The figure represents this as a relationship without a direct link, but with the actions having an impact in the task environment.

One of the limitations of this simple feedback-loop model is that it does not clearly indicate how information is processed by the individual. For most tasks related to command and control, the cognitive processing of information has important influences on actions and eventual performance. However, it is possible to illustrate the role that information processing would play in this model of goal-directed behavior. First, however, we need to indicate what is meant by

information processing. Figure 2 presents a generic information processing model (Hinsz, Tindale, & Vollrath, 1997). The generic information processing model is meant to represent the cognitive processes that may be involved in dynamic operational environments. This model is not the quintessential model of information processing, but rather is one representation that helps us understand how information processing may be involved in these operational environments.

Figure 2
A generic information-processing model.



There are a number of similarities between this generic information processing model and the simple feedback-loop model. Note that both begin with information that is acquired in the context of a goal or objective. However, the processing objective also has an impact on other aspects of information processing. Moreover, both models end with a response or action. Preceding the action, both models imply that the information is used to reach a decision of how to act. In addition, both models involve a component to indicate that the actions influence the environment in a way that contributes feedback as a source of information.

The generic information processing model is more elaborate than the simple feedback-loop model in terms of the cognitive processes related to processing of the information. This generic model suggests that information must be attended to in a meaningful way for it to be processed further. The processing work space indicates that the information is generally considered in the context of some pre-existing knowledge (e.g., working memory). This information is retrieved from an information storage facility. The information in the processing work space can also be considered for possible contribution to the store of knowledge. The information from the processing work space can be examined for its ability to provide meaning,

structure, and organization of the information (encoding), which can then ultimately be placed into storage. The encoding, storage, and retrieval aspects are generally referred to as the memory functions.

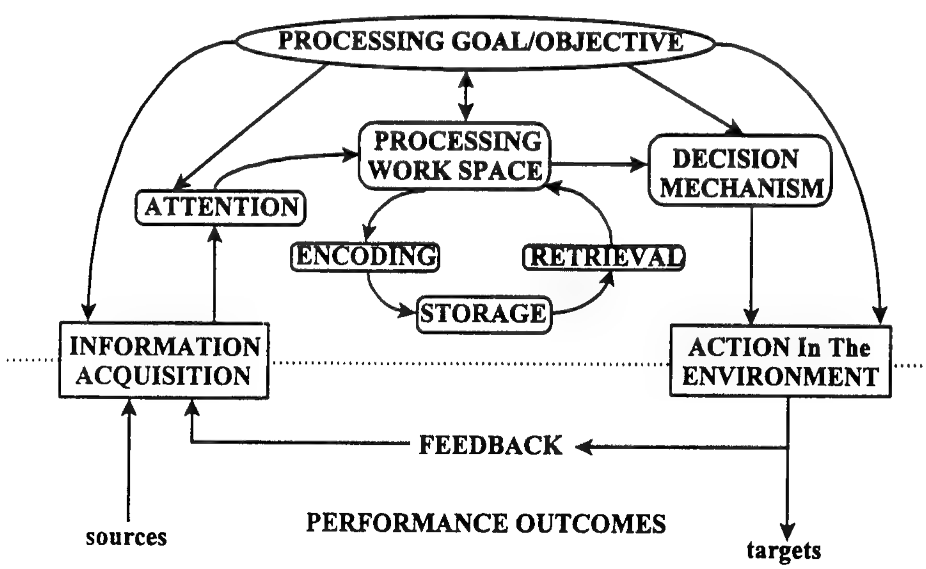
This generic information processing model also reflects the dynamic nature of information processing. The information is considered in relation to prior knowledge and the surrounding context, and the response provides information that feeds back as another source of information for the system. Consequently, information processing can be considered as a subsystem for a larger system of human performance.

It is possible to produce an integrated model that represents the information-processing and goal-directed behavior aspects of performance in a dynamic operational environment (Figure 3). This integrated model combines the similar functions in single components such that both the information processing and simple-feedback loop aspects can be clearly seen and related.

In this representation, multiple sources of information and multiple targets of action are identified. The actions of an individual on a dynamic task can be directed toward a host of targets (e.g., coworkers, technological devices). Moreover, the information used in performing the task can come from a variety of sources (e.g., technological displays, coworkers). The dotted line in this figure can be considered to represent the interface between the human task performer and a technological system used to perform the task.

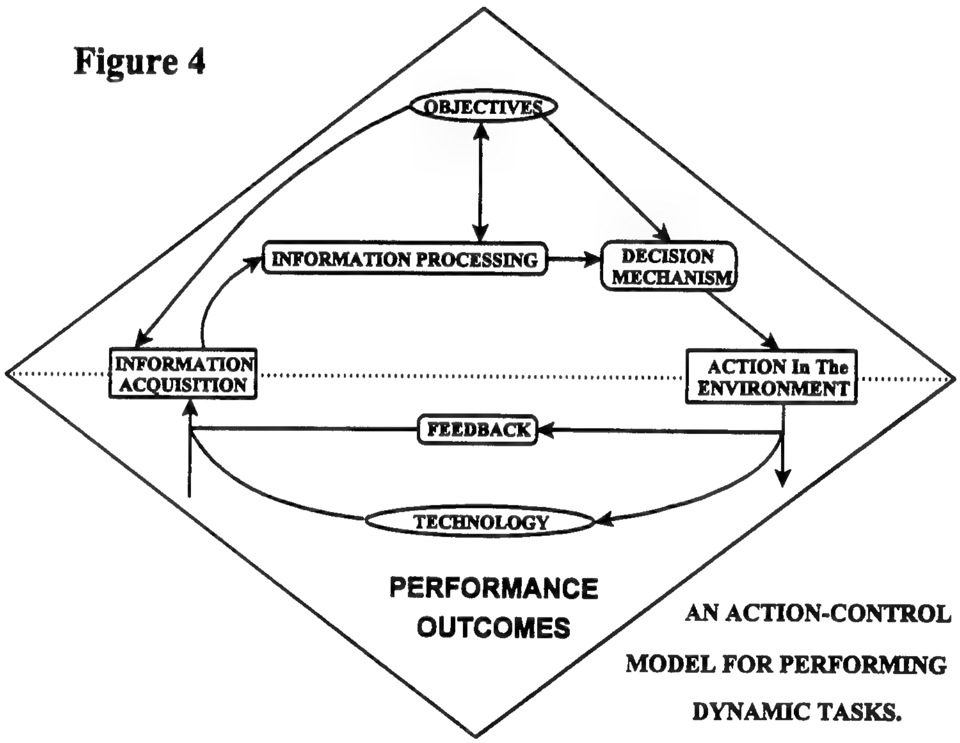
Figure 3

An integrated information-processing and feedback-loop model for actions performed on a dynamic task.



The specific aspects of processing information will not play a primary role in the research being proposed, so a simplified action-control model of performance on dynamic tasks is presented in Figure 4. This figure indicates the processing of information as one component of an overall model that examines the action of the individual in a technological environment. The objectives and information acquired also play a role in controlling the actions of the individual on the task, and the actions are determined by a decision mechanism. This figure also now indicates that outcomes arise from the individual's actions in the environment. Specifically, these would be outcomes associated with performing the dynamic task, which may be intended or unintended outcomes. Figure 4 will serve as the representation of individual performance (termed the action-control model) upon which a model of crew performance will be developed.

Figure 4



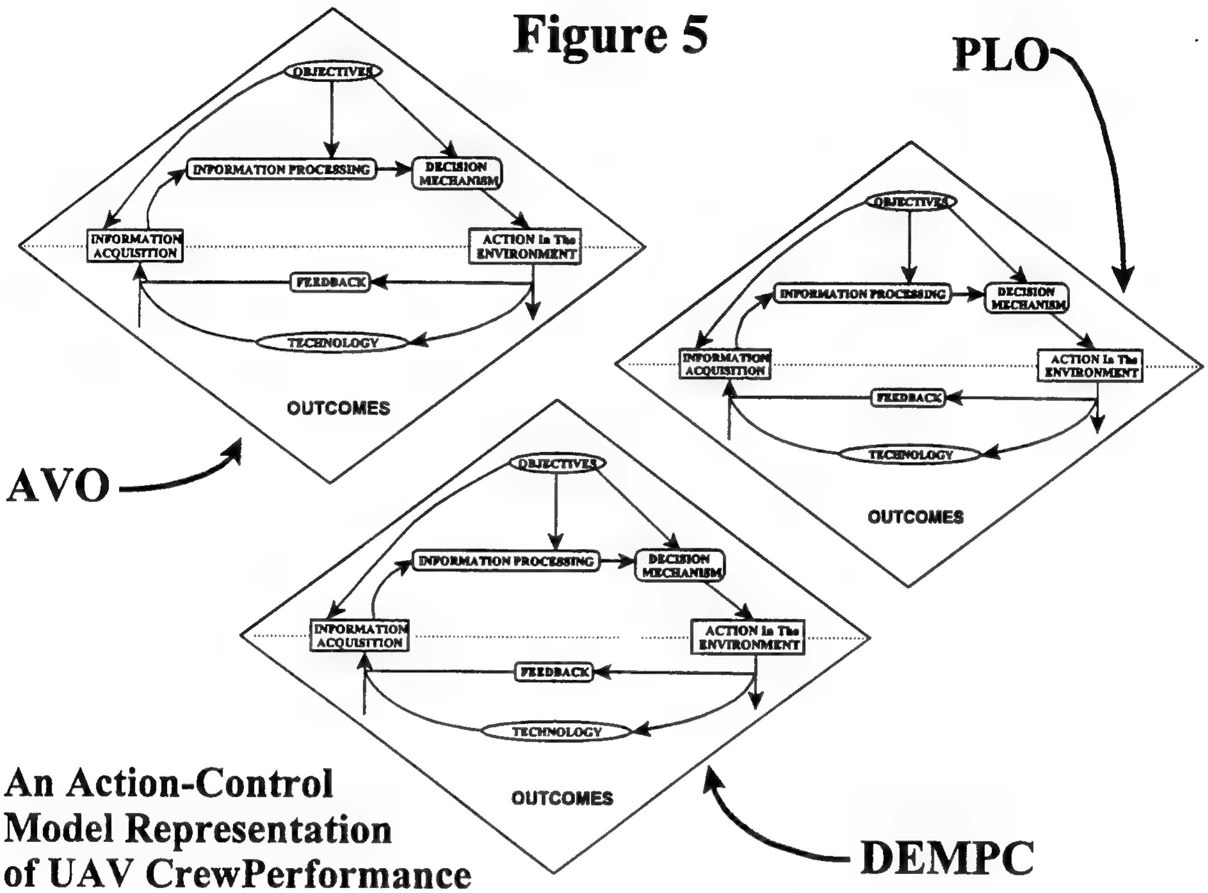
AN ACTION-CONTROL MODEL FOR PERFORMING DYNAMIC TASKS.

Action-Control in a UAV Operational Environment

To place the model of individual performance on a dynamic task in a more specific and relevant context, let us consider the model as it reflects the actions of the individual crew members of a Uninhabited Aerial Vehicle (UAV; see Figure 5). According to a task analysis of the Predator UAV (Hall & Gugerty, 1997), there are three conceptual functions or roles that individual crew members perform to operate the UAV. The air vehicle operator (AVO-piloting) is the individual that flies the vehicle. The payload operator (PLO-photographing) is responsible

for the acquisition of data from the target, in particular, the PLO operates the cameras. The data exploitation and mission planning person (DEMPC-directing) is responsible for planning the mission, and with exchanging information with the operations centers (Ops & RED cells). My illustration will be based on my understanding of the tasks as represented in the task analysis.

Figure 5



An Action-Control Model Representation of UAV CrewPerformance

Let us presume that the UAV is on a mission to acquire data in a non-hostile area. Each member of the crew will have his own objectives for his performance of his tasks and functions (UAV crew members appear to be predominately male and so I will use the male pronoun). The DEMPC will have sketched out the schedule of targets and briefed the AVO and PLO about mission parameters. The DEMPC will monitor how well each target is acquired (in reference to criteria specified by the RED cells). Given his understanding of the mission, the AVO will then attempt to pilot the UAV efficiently so that it safely maintains the appropriate flight path within the specific territory and air space constraints. Because the mission is largely defined by the

targets to be acquired, the AVO also has to arrange the flight path so that the PLO can get optimal photographs of the target. Consequently, the PLO is dependent upon the AVO to get his equipment to a proper heading to take optimal photographs, and is also dependent on the DEMPC for specification of the data to be acquired. Likewise, the DEMPC is dependent upon the PLO to provide optimal photographs so that the mission can progress. Otherwise, the RED cell may demand that additional photographs be taken which would make the schedule of targets infeasible. Moreover, the DEMPC relies upon the AVO to inform him if there are constraints that will change the schedule of targets (i.e., strong head winds, potential cloud obstruction). Consequently, the UAV operational environment has three roles that are highly interdependent for the crew to achieve its mission.

The UAV operational environment is also very dynamic. Changes in wind speed and direction, or cloud cover can interfere with a planned schedule for acquiring the targets. The RED cell can ask for more information about a target, which would require another pass over the target, perhaps at a different angle or from a different direction. The task analysis also suggests that the UAV crew can be asked during its mission to acquire photos of new targets introduced while the UAV crew is performing its mission. These new ad-hoc targets will often require the DEMPC to adapt the schedule of targets, as well as activate changes in the objectives of the AVO and PLO. The task analysis suggests that the UAV task environment is very dynamic in nature, and requires the crew members to adapt to a number of changes in the environment.

This simple description of the UAV operational environment suggests that the crew members performing the tasks are highly interdependent, and thus must coordinate their actions and should have effective collaboration strategies to optimally and successfully complete their missions (USAF Scientific Advisory Board, 1996, Appendix 6D). The functions of each of the members of the crew (AVO, PLO, DEMPC) can each be represented by an action-control model (Figure 4) in one independent arena (Figure 5). However, it is most important to realize that the interdependent nature of the operators tasks suggests that the representation in Figure 5 is inadequate to illustrate the ways in which the members will work together as a crew to complete the mission. The interdependent nature of the AVO, PLO and DEMPC functions is better represented in Figure 6.

A Model of Crew Member Interactions in a Dynamic Operational environment

The hexagon in Figure 6 is divided into three diamonds that each represent one of the functions in the UAV crew: AVO, PLO, and DEMPC. [The representation in Figure 6 is also applicable to other dynamic operational environments in which crew members interact.] Each of the diamonds intersects with the other two diamonds to demonstrate that there is an interrelation between the activities of one with the others. Moreover, the space outside of the hexagon inside the rectangle is meant to represent the environment and context in which the UAV crew performs it's functions. This model provides a basis upon which to consider how the crew members interact with each other and their operational environments in pursuit of the mission objectives.

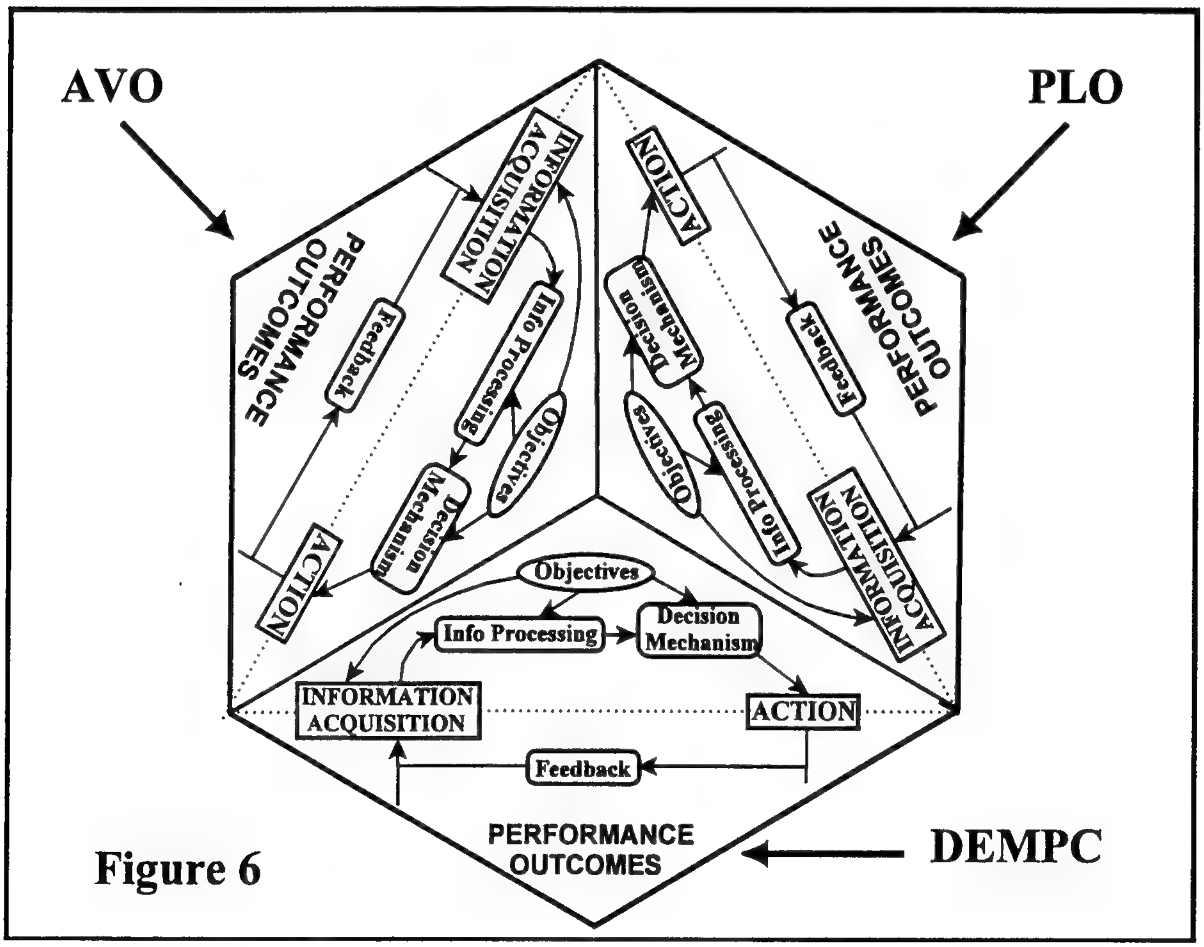


Figure 6

The model of crew member interactions can also be expanded to indicate the flow of information and interactions. Figure 7 provides representations of the crew members encircled by a coordination network (McGrath, et al., 1999). This coordination network demonstrates that

the actions of one member of the crew will pass on to perhaps influence the other members of the crew. Similarly, the actions of others can serve as important information for one of the crew members in pursuit of his particular task objectives. Consequently, the coordination network can represent the "flow" of information and interactions among the crew members.

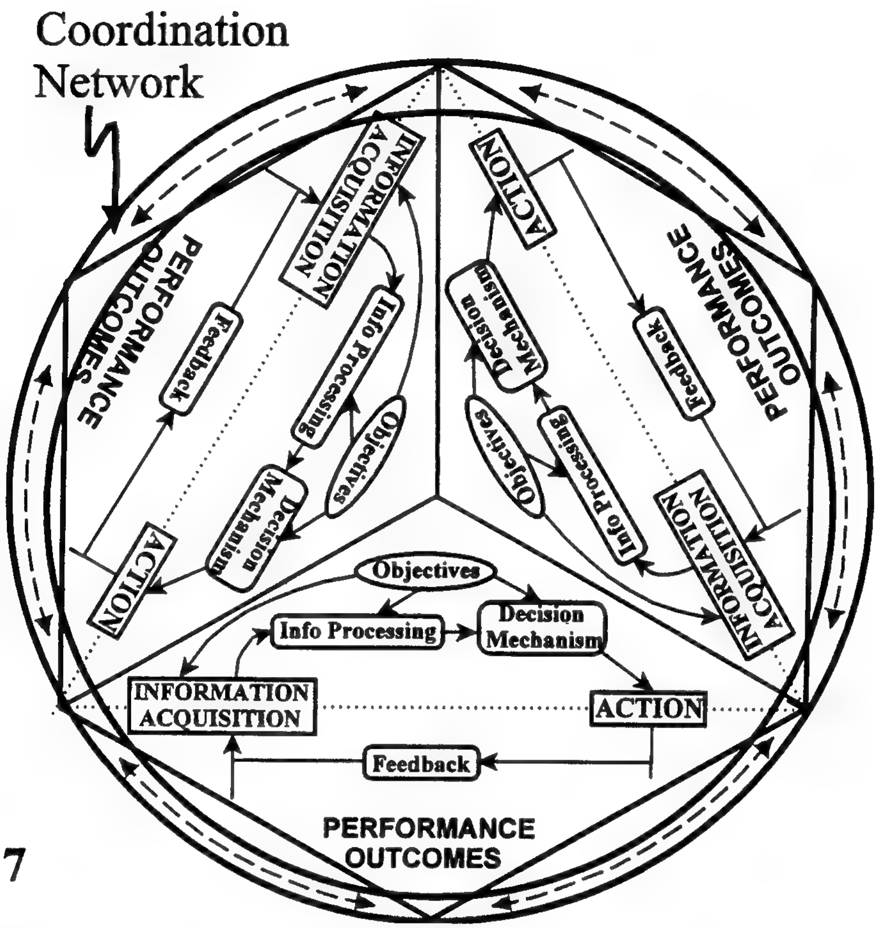


Figure 7

Although the illustration of this model of crew member interactions is two dimensional in nature, one can also envision links between aspects of the functions of each of the crew members as arcs (in the third dimension) between the member action-control models (diamonds). For instance, the feedback one member receives from his actions may be directly relevant to the objectives of another member. The strategies involved in collaboration among the crew members may be represented by a collection of arcs among the crew members.

Up to this point, the models of crew performance have been developed at the individual level of analysis. However, the conceptualization being developed here also allows us to

AN ACTION-CONTROL MODEL FOR CREWS PERFORMING DYNAMIC TASKS

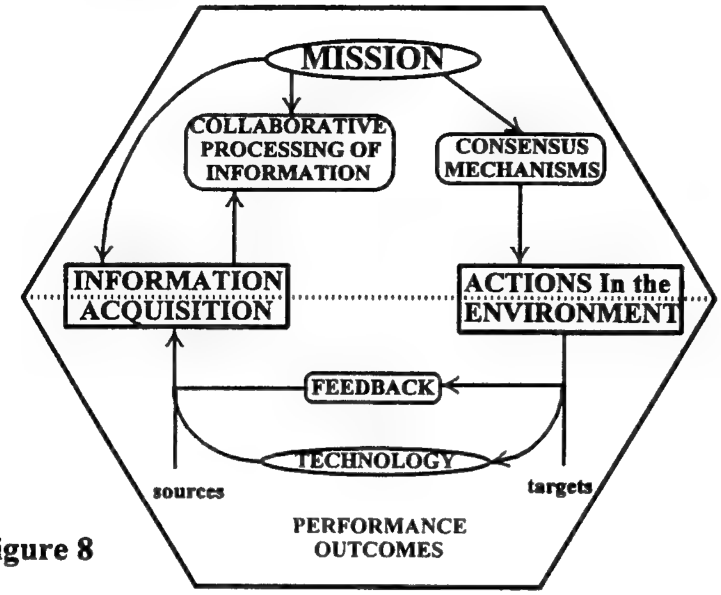
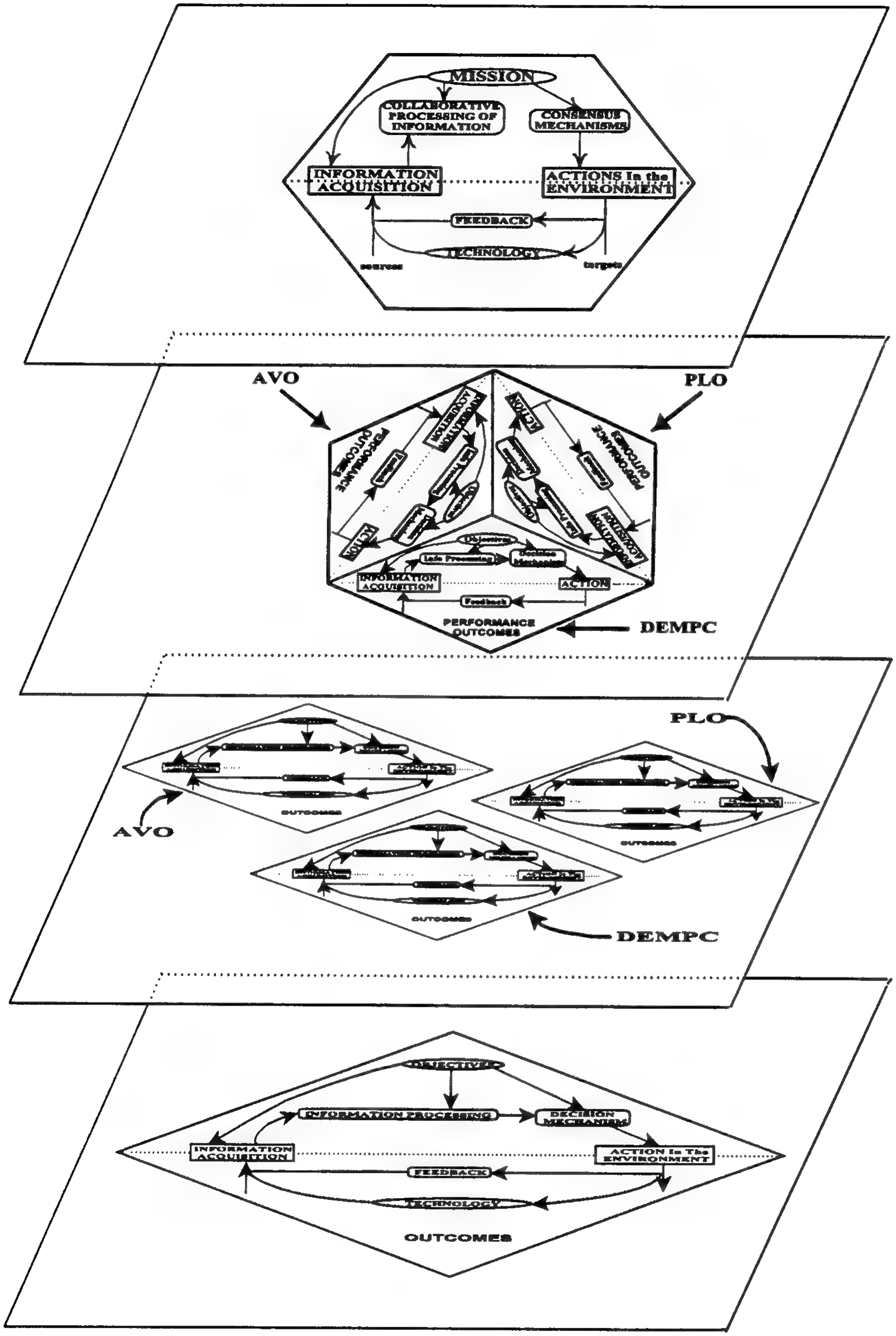


Figure 8

consider a UAV crew as an entity itself. Figure 8 demonstrates that we can conceive of the processes the UAV crew performs as it pursues its mission as an action-control model as well. In this case, the over-arching mission of the crew is specified instead of the objectives of each of the crew members (although they will be interrelated). Moreover, the information processing at this level reflects the ways in which the information is processed collaboratively. In addition, the decision mechanism here reflects the consensus that is achieved about how the tasks will be conducted and the strategies involved. This action-control model of crew performance uses the crew as the level of analysis, and so can also focus on crew-level performance outcomes in addition to the crew member measures of performance.

The general action-control model of task performance was used at both the crew-member and crew-in-entirety levels. This symmetry is intentional so that the hierarchical nature of performance in the dynamic operational environment is demonstrated. Figure 9 demonstrates that crew performance in a UAV operational environment can be represented at multiple levels. The lowest level (local) is that of the crew member performing his specific functions, independent of the activities of the other crew members. The next highest level (local collective) indicates that the different crew members can be seen as performing their tasks in the same environment. The second from highest (local integrated) level illustrates the interactions that occur among the crew members as they perform their tasks. Coordination and collaboration

Figure 9


MISSION
COLLABORATIVE PROCESSING OF INFORMATION
CONSENSUS MECHANISMS
INFORMATION ACQUISITION
ACTIONS in the ENVIRONMENT
FEEDBACK
TECHNOLOGY
sources
targets
AVO
PLO
PERFORMANCE OUTCOMES
Objectives
Info Processing
Decision Mechanism
INFORMATION ACQUISITION
ACTION
Feedback
PERFORMANCE OUTCOMES
DEMPC
PLO
AVO
DEMPC
OBJECTIVES
INFORMATION PROCESSING
DECISION MECHANISM
INFORMATION ACQUISITION
ACTION in The ENVIRONMENT
FEEDBACK
TECHNOLOGY
OUTCOMES

among the crew members are the focus of this level. Finally, the top level (global) represents the actions of the crew as a functioning unit as it pursues its mission. Although not presented in Figure 9, it is implied that there would be a host of relations between the different levels in this hierarchy. That is, that the activities at the crew level would influence the actions in the interaction level that may influence the action at the crew member level. Likewise, relations that represent upward influence would also arise. The actions of the crew member on his tasks would influence the interactions among the crew members that may have impact on the activities of the crew as a whole.

This series of figures is intended to illustrate a conceptualization of crew performance in a dynamic, complex operational environment such as the UAV. This conceptualization suggests that performance of a crew in a UAV operational environment may be considered as a set of embedded subsystems which form a hierarchy of levels of analysis. This series of figures and the models they represent illustrate the complexity of the nature of dynamic operational environments. Considering these illustrations, it is possible to elaborate a large number of research questions that could drive efforts to understand crew performance in a UAV operational environment.

Research Agendas for the Study of Crew Performance in a UAV Operational Environment

Figures 6 and 7 present representations of models of how crew members interact and perform their functions. These models can also provide a basis upon which to develop a set of research agendas. A research agenda could be established for each of the aspects of the models (e.g., coordination network, feedback). As one example, at the center of the hexagon representing the interactions among the crew members, the objectives of each member of the crew are reflected. The objectives and goals of the crew members for performing their tasks could be a focus of research. Some questions that might be raised concern how potential conflicts among the objectives of the separate crew members might be resolved. Moreover, it would be possible to examine how the interdependence in terms of goals and objectives can be constructed to support the most effective performance. The existing literature on goal interdependence (cf., O'Leary-Kelly, Martocchio, & Frink, 1994; Mitchell & Silver, 1990) could serve as a foundation upon which to develop strategies that would reduce goal conflict among the

interdependent crew members, while at the same time use goals to enhance the performance of the crew members.

If the three diamonds in the hexagons (used to represent the model of crew interaction and performance in Figures 6 and 7) are imagined to rotate slightly, the decision mechanism aspect of the action-control model could become the center of attention. A research agenda can be developed that examines the processes by which the crew decisions are reached about the cognitive and behavioral strategies to be used. Additionally, research questions about how the behavioral and cognitive strategies are to be integrated to direct the actions of the crew members can be examined.

The decision mechanism can represent the members' intention to act (Fishbein & Ajzen, 1975; Hinsz & Ployhart, 1998). As such, these intentions will play a large role in actual collaboration and coordination of the crew. How do the crew members come to agree on and make trade-offs regarding their plans and strategies? A number of different research programs arise from considering the inter-relations among the crew members' task performance decisions.

Another research program could address the impact of mental models of crew performance on the actions that result. The mental models in this domain would also relate to situation awareness, and more interestingly, shared situation awareness (Endsley, 1995). These shared mental models can influence a common operational understanding of the situation by crew members. The role of shared mental models in team performance has received considerable discussion (Cannon-Bowers, Salas, & Converse, 1993; Endsley, 1995; Hinsz, 1995a; Hinsz, 1996), however, the actual empirical examination of the potential of shared mental models is much more meager.

Another research program could examine the impact of structuring arrangements among the crew members for developing task performance strategies. A task analysis of AWACS weapons directors (Fahey, Rowe, Dunlap, & deBoom, 1997) suggests that members of the AWACS crew meet before the start of a mission to negotiate arrangements (contracts) about how tasks and responsibilities will be shared and coordinated. These contracts appear to be important features of the AWACS crews, and likely contribute to the crew effectiveness. Research on teams suggests that teams rarely structure their interactions (Hackman, 1987), even though teams

are known to abhor a vacuum (Hinsz, 1995b). One potential means of enhancing the performance of UAV crews would be to establish a structure in which members of the crew would meet to negotiate arrangements regarding how tasks and responsibilities would be coordinated. These negotiated arrangements might enhance shared situation awareness, as well as contribute to better strategies for teamwork (cf., Rentsch & Hall, 1984).

If the three member components (diamonds) of the model of crew interaction (Figures 6 & 7) are again rotated slightly, the actions of each of the crew members would be the center of attention. Research highlighted by this situation would focus on how the behaviors of each crew member could facilitate or hinder the activities of the other crew members. It is at this point that we could also consider the impact of technological aids (e.g., adaptive interfaces) to enhance the actions of the crew members in the performance of their tasks.

If each of the components (diamonds) of the model of crew interaction are rotated entirely 180 degrees from the way they are represented in Figures 6 and 7, the role of feedback on task performance could be the focus of research. Although feedback to individual task performers has received considerable attention, much less research has addressed collective or crew feedback (Hinsz, et al., 1997; Robinson & Weldon, 1993). One interesting and important question that could be addressed is the amount of feedback provided the crew members regarding individual task performance and the crew's overall performance on the task. Moreover, should each member of the crew receive feedback about the consequences of the other crew members' actions (e.g., should the AVO learn of the quality of photographs taken by the PLO), or will this produce information overload, or even produce divisions among the crew members.

A further rotation of member components of the model of crew interaction would place the acquisition of information as the center of attention. Conflicts among the crew members of a UAV may arise because of different objectives and finite resources. In UAV missions that are beyond line of sight, the bandwidth of the signal is limited. Consequently, images from either the payload camera, or a nose camera of the UAV can be viewed. However, images from both cameras can not be viewed simultaneously. Consequently, if the PLO is using the payload camera to arrange a sequence of shots, the nose camera is not available for the AVO's use. The AVO may be concerned about flying around an obstacle (e.g., mountain) and may desire to use

the nose camera. Because both cameras can not be used simultaneously, the AVO and PLO must arrange to share the resource. In this case, they would have to coordinate efforts so that both can acquire information that is important for their own individual objectives.

As mentioned earlier, information processing per se will not be a central focus of this paper. However, a number of questions would arise if the information processing of crew members in pursuit of the mission is the focus of research attention. For instance, the DEMPC will influence the information that might receive attention by the AVO and PLO. The memory processes of one member might contribute to a transactive memory system (Wegner, 1987) so that he could inform another member about a particular problem. For example, one member might recall that on a previous mission, a similar maneuver was required to acquire a target. Much like recognition-primed decision making (G. Klein, 1989), this member could then inform the other about the situation and the effective means of dealing with it.

Another way in which information processing could be involved is in terms of biases that might be prevalent among the crew members. A shared bias to perceive a specific situation in an ineffective fashion (e.g., We can't do this.), could mean that more in-depth analysis of the situation is not attempted, and the crew fails to completely fulfill its mission. Research on information processing biases in groups suggests that the bias is often exaggerated in a group response (Hinsz, et al., 1997; Tindale, 1993).

This overview of research agendas is not intended to list all the potential research projects that might be conducted regarding crew performance. Rather, this review hopefully serves to demonstrate how the conceptualization provided helps to identify a host of research agendas that can be pursued. One feature of providing a conceptual model of crew performance in a dynamic operational environment is that it integrates a number of ideas from a host of domains to provide a representation that helps us understand the processes involved. Based on the conceptualization provided, it is now possible to see how a number of research projects could all contribute to the expansion of our understanding of the processes involved in crews performing in the UAV and other similarly complex operational environments.

References

Cannon-Bowers, J.A., Salas, E., & Converse, S.A. (1993). Shared mental models in team decision making. In N.J. Castellan, Jr., (Ed.), Individual and group decision making (pp. 221-246). Hillsdale, NJ: Erlbaum.

Carver, C.S., & Scheier, M.F. (1981). Attention and self-regulation: A control theory approach to human behavior. New York: Springer-Verlag.

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37, 32-64.

Fahey, R.P., Rowe, A.L., Dunlap, K.L., & deBoom, D.O. (1997). Synthetic task design I: Preliminary cognitive task analysis of AWACS weapons director teams. Report for the Air Force Research Laboratory (HEJC), Brooks AFB, Texas.

Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention, and behavior: An introduction to theory and research. Reading, MA: Addision-Wesley.

Hall, E., & Gugerty, L. (1997). Predator operations cognitive task analysis results. Briefing prepared by the Air Force Research Laboratory, Brooks AFB, Texas.

Hackman, J.R. (1987). The design of work teams. In J.W. Lorsch (Ed.), Handbook of Organizational Behavior (pp. 315-342). Englewood Cliffs, NJ: Prentice-Hall.

Hinsz, V.B. (1995a). Mental models of groups as social systems: Considerations of specification and assessment. Small Group Research, 26, 200-233.

Hinsz, V.B. (1995b). Group and individual decision making for task performance goals: Processes in the establishment of goals in groups. Journal of Applied Social Psychology, 25, 353-370.

Hinsz, V.B. (1996). Metacognition in groups and the potential of shared mental models. Paper presented at the Fourth Annual Society for Experimental Social Psychology Preconference on Small Groups, Sturbridge, MA.

Hinsz, V.B., & Ployhart, R.E. (1998). Trying, intentions, and the processes by which goals influence performance: An empirical test of the theory of goal pursuit. Journal of Applied Social Psychology, 28, 1050-1065.

Hinsz, V.B., Tindale, R.S., & Vollrath, D.A. (1997). The emerging conceptualization of groups as information processors. Psychological Bulletin, 121, 43-64.

Kanfer, R. (1992). Work motivation: New directions in theory and research. In C.L. Cooper & I.T. Robertson (Eds), International Review of Industrial and Organizational

Psychology (Vol. 7, pp. 1-53). London: John Wiley & Sons.

Klein, G.A. (1989). Recognition-primed decisions. Advances in Man-Machine Systems Research, 5, 47-92.

Klein, H. J. (1989). An integrated control theory model of work motivation. Academy of Management Review, 14, 150-172.

Lord, R.G., & Kernan, M.C. (1989). Application of control theory to work settings. In W.A. Herschberger (Ed.), Volitional action (pp. 493-514). Amsterdam: Elsevier.

McGrath, J.E., Arrow, H., & Berdahl, J.L. (1999). A theory of groups as complex systems. Newbury Park, CA: Sage.

Mitchell, T.R., & Silver, W.S. (1990). Individual and group goals when workers are interdependent: Effects on task strategies and performance. Journal of Applied Psychology, 75, 185-193.

O'Leary-Kelly, A.M., Martocchio, J.J., & Frink, D.D. (1994). A review of the influence of group goals on group performance. Academy of Management Journal, 37, 1285-1301.

Powers, W.T. (1973). Feedback: Beyond behaviorism. Science, 179, 351-356.

Rassmussen, J., Pejtersen, A.M., & Goodstein, L.P. (1994). Cognitive systems engineering. New York: Wiley.

Rentsch, J.R., & Hall, R.J. (1994). Members of great teams think alike: A model of team effectiveness and schema similarity among team members. Advances in Interdisciplinary Studies of Work Teams, 1, 223-261.

Robinson, S., & Weldon, E. (1993). Feedback seeking in groups: A theoretical perspective. British Journal of Social Psychology, 32, 71-86.

Smith, K.C.S. (1996). Decision making in rapidly changing environments: Trading in the spot currency markets. Doctoral Dissertation, University of Minnesota.

Tindale, R.S. (1993). Decision errors made by individuals and groups. In N.J. Castellan, Jr. (Ed.), Individual and group decision making (pp. 109-124). Hillsdale, NJ: Erlbaum.

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G.R. Goethals (Eds.), Theories of group behavior (pp. 185-208). New York: Springer-Verlag.

USAF Scientific Advisory Board (1996). UAV technologies and combat operations. US Air Force Scientific Advisory Board, Pentagon, Washington, D.C.

STATISTICAL MODELS FOR ALTITUDE DECOMPRESSION SICKNESS

Nandini Kannan
Associate Professor
Division of Mathematics and Statistics

University of Texas at San Antonio
6900 N. Loop 1604 West
San Antonio, TX 78249

Final Report for:
Summer Faculty Research Program
Armstrong Research Site
AFRL/ HEPR, Brooks Air Force Base

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Research Site
AFRL/ HEPR, Brooks Air Force Base

September 1998

STATISTICAL MODELS FOR ALTITUDE DECOMPRESSION SICKNESS

Nandini Kannan
Associate Professor
Division of Mathematics and Statistics
University of Texas at San Antonio

Abstract

Survival analysis techniques were used to model the phenomenon of altitude decompression sickness. Models based on the loglogistic distribution were developed to predict the probability/ risk of DCS over time. Several risk factors such as exercise, pressure, and preoxygenation times were included in the model to assess their effects on DCS. To account for the effects of exercise at high altitudes, the altitude range was stratified and separate models developed for the different strata. The effects of exercise during preoxygenation were also studied. Using experimental data, maximum likelihood methods were used to estimate the model parameters. The fitted model was then used to predict the probability of DCS for several exposure profiles. These predictions agreed closely with the empirical DCS percentages from the database.

STATISTICAL MODELS FOR ALTITUDE DECOMPRESSION SICKNESS

Nandini Kannan

Introduction

Symptoms of decompression sickness (DCS) may occur when individuals are exposed to significant change in environmental pressure. These changes may occur during diving, high altitude exposures or artificially induced pressure changes in hyperbaric or hypobaric chambers. Symptoms of DCS include joint pain, respiratory difficulties, central nervous system problems, and in extreme situations, even death. The cause of these manifestations has been a subject of extensive research over the years. Physiologists agree that the evolution of nitrogen bubbles in body tissues occurring as a result of changes in ambient pressure cause the symptoms. For significant changes in pressure, the inability of tissue gas exchange processes to expel excess nitrogen results in supersaturation of the tissues and blood. These gases which come out of solution when tissues are sufficiently supersaturated collect as bubbles in the tissue. Their size and location are thought to have a significant effect on the occurrence of DCS symptoms.

Aircrews operating in unpressurized aircraft, aircraft with inadequate pressurization, and extravehicular activity in space must assume some level of DCS risk. The risk depends on the final altitude, length of exposure, preoxygenation time, type of breathing gas, exercise, and many other factors affecting bubble formation and growth. The risk of DCS may be eliminated/ decreased by controlling the levels of these "risk factors". For example, breathing 100 % oxygen prior to ascent helps in removing the nitrogen dissolved in blood and tissues. This, in most situations, will result in a delayed onset of symptoms of decompression sickness. The ability to predict DCS risk real-time, as well as for mission planning, is an operational need for both military and civilian aerospace applications. In order to predict the risk, it is

necessary to understand the effects of the various risk factors and use that information in developing an appropriate model.

To study the occurrence of altitude DCS and its primary risk factors, the Air Force Research Laboratory has for many years been conducting research using human subjects exposed to simulated altitude in hypobaric chambers. The subjects are exposed to a variety of altitudes, exposure times, preoxygenation times, breathing gas mixtures, and levels of exercise. While at simulated altitude, the subjects are monitored for symptoms of DCS and for venous gas emboli (VGE) using echocardiography (Hewlett Packard SONOS 500,1000, and 1500). Data from these experiments have been deposited in the Air Force Research Laboratory DCS Research Database. At present the database contains over 2000 subject records.

I have been involved with this project for over two years. Our primary focus has been the development of a probabilistic model using the database that can predict the probability/ instantaneous risk of developing DCS for different flight profiles. Using survival analysis techniques, an initial model was derived using the loglogistic distribution. At present, I am working on modifications and improvements to the model. The final product of this research effort will be a validated mathematical model that will provide the foundation for development of a standardized altitude decompression computer for DCS risk assessment and management in USAF flight and altitude chamber operations.

Background

Decompression sickness is commonly experienced after extended hyperbaric exposures. Most of the available literature in modeling DCS risk considers diving operations. In contrast, the aerospace field has never had a standardized approach for predicting DCS risk. Diving models cannot adequately predict risks at altitude because of certain characteristics that are unique to high altitude exposure. The first significant difference is the faster ascent rate in altitude exposures compared to the

slower ascent rates for divers. Second, there is a significant threat of instantaneous pressure loss in aircraft. The most significant difference is that symptoms of DCS occur during mission for altitude exposure, in contrast to post mission appearance of symptoms for diving exposures.

Statistical methods in altitude DCS modeling have only recently been applied. Weathersby [1], [2] used the method of maximum likelihood to fit several models to DCS data. This was the first attempt to quantify decompression risk. In the second paper, Weathersby provided probabilistic models for predicting the occurrence of DCS. The article however considered diving data. Van Liew et al. [3] developed a probabilistic model of altitude DCS in which the mechanistic principles used were based on the premise that risk of DCS is related to the number of bubbles and the volume of gas that can be liberated from a unit of tissue. The authors developed equations that incorporated these premises, and used these equations in the risk function. They tested several models and used the model that best fit the data as the most realistic one. The covariates (risk factors) used in the model were duration of 100% O_2 at ground level (preoxygenation), atmospheric pressure after ascent, and exposure duration.

Gerth and Vann [4] developed an extensive model for bubble dynamics to predict DCS using equations similar to those in Van Liew et al. Treating the percentage of individuals with DCS as the response variable, maximum likelihood methods were used to estimate the model parameters. In an addendum, they discussed the importance of including onset times to provide a significant improvement in the predictions.

There are some drawbacks to the above approaches. First, several assumptions have to be made in order for these bubble dynamic equations to hold. There is almost no data available on bubble characteristics other than measures of VGE on a Spencer scale. These models also ignore the time to onset of DCS and do not adequately incorporate information on asymptomatic individuals.

To counter these problems, Kumar et al. [5], [6], [7], [8] in a series of papers,

recognized that survival analysis techniques are the most appropriate to model DCS risk. They developed logistic and loglinear models to predict DCS as a function of Tissue Ratio, which is a measure of tissue nitrogen decompression stress. Another covariate used was CMB (circulating microbubbles) status. The models used the logarithm of time to DCS and maximum likelihood techniques to estimate the model parameters. The articles allude to the fact that censoring occurs for individuals who did not exhibit any symptoms of DCS. Conkin et al [9] in a recent paper have discussed in some detail the use of survival times and censoring using the loglogistic model. They also discuss different forms of the risk functions using certain mechanistic assumptions similar to those of Van Liew and others.

There are certain limitations to the models proposed in these papers. The series of articles examines the effect of one or two factors on the DCS risk. However, the risk of DCS is affected by a number of competing factors like the preoxygenation time, exposure time, exercise status, and altitude. In order to develop a model that adequately describes the phenomenon of DCS, all these factors should be included in a survival model. This would determine the relative importance of the different factors, and possibly provide a method of controlling the risk of DCS. These models also ignore certain aspects of the data like the varying number of subjects in the different protocols, and the large dispersion in the data.

In a recent article, Kannan, Raychaudhuri, and Pilmanis [10] have developed a survival model using the loglogistic distribution. The authors have shown that the most significant risk factors are final altitude, ratio of preoxygenation time to exposure time, and level of exercise during the exposure. Because of the large dispersion in the data, weights were introduced in the model. Validation and cross-validation techniques were used to evaluate the predictions from this model. For a wide spectrum of exposure profiles, the predictions from the model agree closely with empirical probabilities from the database. In a subsequent technical report, Kannan [11] has developed confidence bands for the survival and risk functions using bootstrap tech-

niques. The confidence bands clearly suggest that the loglogistic model is indeed an excellent predictive model for most exposure profiles.

In this report, we will evaluate the importance of certain risk factors in greater detail. The effects of very high altitudes ($\geq$ 30,000 ft) and the relation between altitude and exercise will be studied in detail. Analysis of profiles with exercise during preoxygenation and inflight denitrogenation will also be considered.

Methodology

Survival Analysis is the study of lifetimes or failure times of individuals. In studying altitude DCS, the primary variable of interest is T, the time to onset of symptoms. Since individuals exposed to identical conditions react differently, it is clear that T is a random variable. We can use several distributions to model T including the Weibull, gamma, lognormal etc. These distributions all have their unique functional form and the choice of the appropriate distribution is largely determined by the data.

In survival analysis, one of the most important functions is the hazard or risk function that measures the instantaneous rate of developing symptoms at time t, given the individual is symptom free up to that time. The risk function can be a monotonic (increasing or decreasing) function of time, or it can be nonmonotonic. Empirical evidence has shown that the risk of DCS increases with exposure, reaches a maximum, and then starts to decrease because of denitrogenation. This reduces the class of distributions that are appropriate for T to the loglogistic and the lognormal.

Another characteristic unique to survival analysis is the notion of censoring. For any flight profile, several subjects will not exhibit symptoms during the exposure period; these are the censored observations. These individuals may have developed symptoms if the exposure period were increased. The database has approximately 40 % censoring; i.e. almost 40 % of subjects reported no symptoms during the experiments. It is important that information on these subjects be appropriately incorporated into any model that is used to predict DCS risk.

Once the choice of an appropriate distribution for T has been made, risk factors must be included in the model. The likelihood function which describes the probability of observing the actual data values (including information on the censored observations) is then constructed. The likelihood function depends on several parameters corresponding to the different risk factors and the scale parameter for the loglogistic distribution. This function is then optimized to obtain the maximum likelihood estimates of the unknown parameters. For a detailed description of survival analysis, the reader is referred to the books by Lawless [12] and Lee [13].

Results

Experiments on human subjects exposed to simulated altitudes in hypobaric chambers has been conducted for several years at the Air Force Research Laboratory at Brooks AFB. Both male and female subjects participated in the study. The subjects were all Air Force personnel aged 18-45 with similar physical characteristics, representative of the USAF rated aircrew population. All flight profiles included a preset altitude, exposure time, preoxygenation time, and exercise regimen to be followed. Subjects were monitored continuously for DCS and VGE. The experiment either lasted the entire exposure period or was terminated if the individual had symptoms. For further details on the different profiles, refer to Webb et al [14].

The data collected in the study included the time to onset of DCS symptoms (ONSET), time of preoxygenation (BR), pressure (PRES), planned time at maximum altitude (TALT), and exercise code (EX). For individuals reporting no symptoms, the onset time was replaced by their corresponding TALT times. These are the censored observations, where the censoring is fixed or Type I. A censoring variable (CENSOR) was created to indicate the status of each individual: 1–DCS and 0–No DCS (Censored).

For some studies, the same individuals performed as many as 5 exposures. These studies were designed to test the effects of inflight denitrogenation or the effects

of increasing exercise on DCS. To avoid any possibility of data contamination, we decided to use only results of the first exposures of these individuals. Even though this resulted in a smaller dataset, we felt that this would reduce the effects of susceptible or resistant individuals. The reduced dataset contained 823 observations of which 357 were censored observations, i.e., 43 % of individuals did not exhibit any symptoms of DCS during the duration of the experiment.

Pressure levels in the dataset ranged from 141 mm to 314 mm. The preoxygenation times ranged from 0 to 240 minutes, and the exposure time ranged from 120 minutes to 480 minutes. Different types of exercise were performed by the subjects. They were classified according to the amount of oxygen consumption as rest, mild exercise, and heavy exercise. There are several articles that have analyzed specific aspects of the database, in particular refer to Sulaiman et al [15], Ryles et al [16], and Webb et al [17].

The earlier model developed by Kannan et al [10] considered three risk factors: PRES, BRTALT (ratio of preoxygenation to exposure time), and EX. In most experiments, we found that individuals who prebreathed longer were more likely to be subject to higher altitudes or longer exposure periods. The BRTALT variable was created to remove this bias. Validation and cross validation techniques were used to conclude that predictions from this model agreed closely with empirical data for most exposure profiles.

The model, however, did not perform satisfactorily for low preoxygenation times, and at high altitudes. The classification of exercise as rest, mild, and heavy did not seem adequate, especially at the higher altitudes. Because of the paucity of data on heavy exercise at present, it would seem reasonable to consider only two cases : rest or exercise. At very high altitudes (30,000 +), the effects of even mild exercise are very pronounced. At 30,000 ft with mild exercise, almost 80 % of subjects bent, compared to 52 % at 29,500 ft. For the resting profile at 30,000 ft, 53 % of subjects reported symptoms. This seems to suggest either a strong interaction effect between

altitude and exercise or some sort of threshold at 30,000 feet. It has been observed in the literature that there is a lower threshold for altitude DCS, i.e. below 21,000 ft very few cases of DCS occur. The risk increases sharply at altitudes above this threshold. It seems reasonable to assume such a threshold may exist at the higher altitudes.

To address the concerns outlined above, it was decided to fit separate models for three different altitude intervals. The three strata were chosen after a careful examination of the data. The first strata consists of pressure levels in the range (282 , 314], i.e. altitudes between 22,500 and 25,000 ft. The second strata consists of pressure levels in the range (226, 282], i.e. altitudes between 25,000 and 30,000 ft. The final strata consists of pressure levels in the range [179, 226], i.e altitudes above and including 30,000 ft. In each of the strata, the exercise variable was given different values and the interaction between pressure and exercise level included as a risk factor.

The model used to fit the data was based on the loglogistic distribution, with survival function given by

$$S(t) \quad = \quad \frac{1}{1 + (\lambda * t)^{\gamma}}$$

The cumulative distribution function (Cdf) is defined as $F(t) = 1 - S(t)$, i.e. the probability of developing symptoms by time t.

The risk function for the loglogistic is given by

$$r(t) \quad = \quad \frac{\lambda\gamma(\lambda * t)^{(\gamma-1)}}{1 + (\lambda * t)^{\gamma}}.$$

The parameter $\gamma = 1/\sigma$, where σ is a scale parameter. If $\sigma < 1$, the risk function of the loglogistic increases to a peak, and then decreases towards zero. This is the shape which accurately describes the risk of DCS over time.

The parameter λ depends on the vector of risk factors $\boldsymbol{x}$ through the following

Table 1: Parameter estimates: Altitude < 25,000 ft

Variable	DF	Estimate	Std. Err.	Chi-sq	p-value
INT	1	25.93	51.39	0.25	0.6139
PRES	1	-2.96	8.91	0.11	0.7396
BR	1	0.009	0.006	1.88	0.1700
EXPR	1	-0.29	0.09	9.62	0.0019
SCALE	1	0.74	0.10		

equation

$$\lambda = exp(-\boldsymbol{\beta}'\boldsymbol{x})$$

where $\boldsymbol{\beta}$ is a vector of unknown parameters which will be estimated from the data.

Using the functions defined above, we can write the likelihood function

$$L(\boldsymbol{\beta}, \gamma) = \prod_{i=1}^{M} f(t_i) \prod_{j=1}^{N-M} S(t_j)$$

where M is the number of uncensored observations. Here $f(t) = F'(t)$ is the probability density function. We use the statistical software package SAS to maximize the likelihood and obtain estimates of the unknown parameters. The results are provided in Tables 1-3. Note that for all three strata, the scale parameter is less than one. The table provides the estimates of the parameters, the standard error of the estimates, and a chi-square value used to assess the relative importance of the different risk factors.

For Strata 1, the interaction between exercise and pressure is the most significant risk factor. Pressure by itself is not that significant, which is not surprising considering the lower altitudes in this range. For Strata 2, both preoxygenation time, and the exercise-pressure interaction are very significant.

For Strata 3, all the factors are highly significant. The exercise-pressure interaction has the highest chi-square value, and is clearly the factor that has the most significant

Table 2: Parameter estimates: 25,000 ft $\leq$ Altitude < 30,000 ft

Variable	DF	Estimate	Std. Err.	Chi-sq	p-value
INT	1	3.997	4.262	0.88	0.3483
PRES	1	0.511	0.775	0.43	0.5105
BR	1	0.008	0.002	11.6	0.0007
EXPR	1	-0.18	0.039	20.53	0.0001
SCALE	1	0.53	0.04		

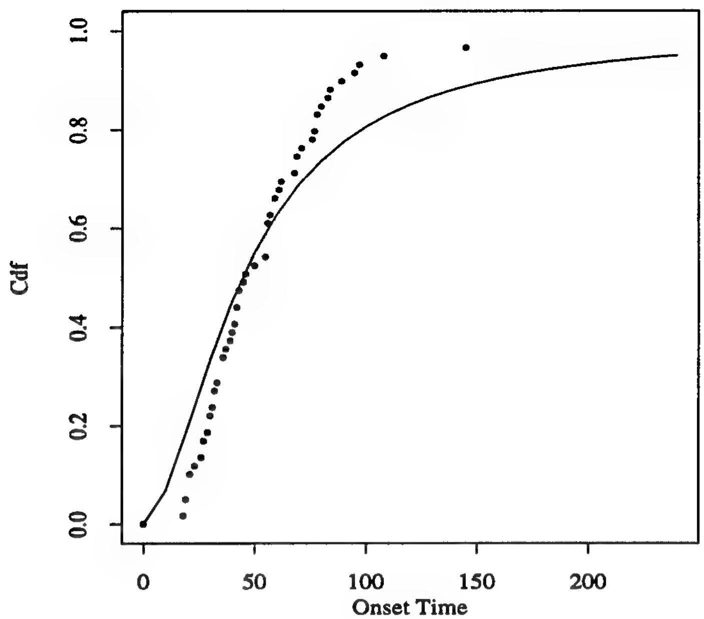
Table 3: Parameter estimates: Altitude $\geq$ 30,000 ft

Variable	DF	Estimate	Std. Err.	Chi-sq	p-value
INT	1	-6.529	2.531	6.66	0.0099
PRES	1	2.613	0.480	29.61	0.0001
BR	1	0.003	0.001	15.61	0.0001
EXPR	1	-0.217	0.029	52.38	0.0001
SCALE	1	0.565	0.029		

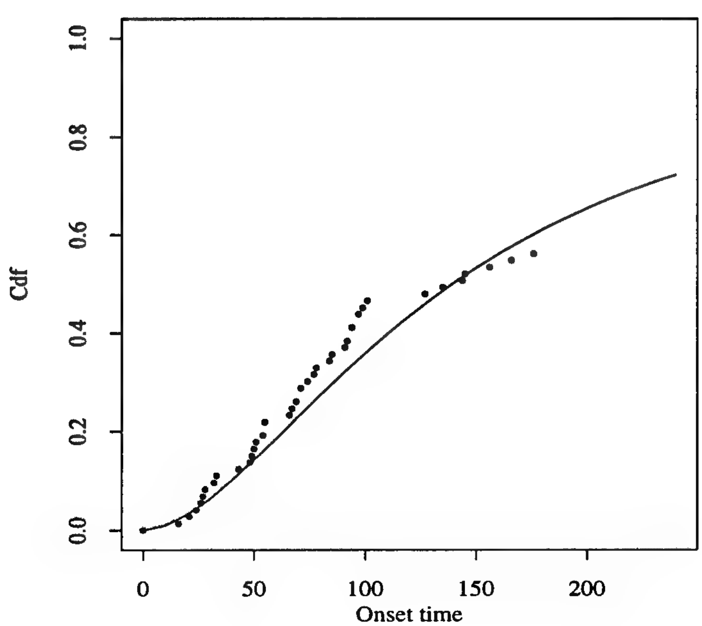
effect in increasing the probability of DCS.

In order to evaluate the predictions from the model, we use the estimates obtained from SAS to predict the probability of DCS over time for several exposure profiles. The graphs are given below in Figures a-e. The solid line is the predicted probability of DCS over time using the loglogistic model. The dots represent the empirical probabilities from the database. We have selected profiles from all three strata to show the versatility of the model.

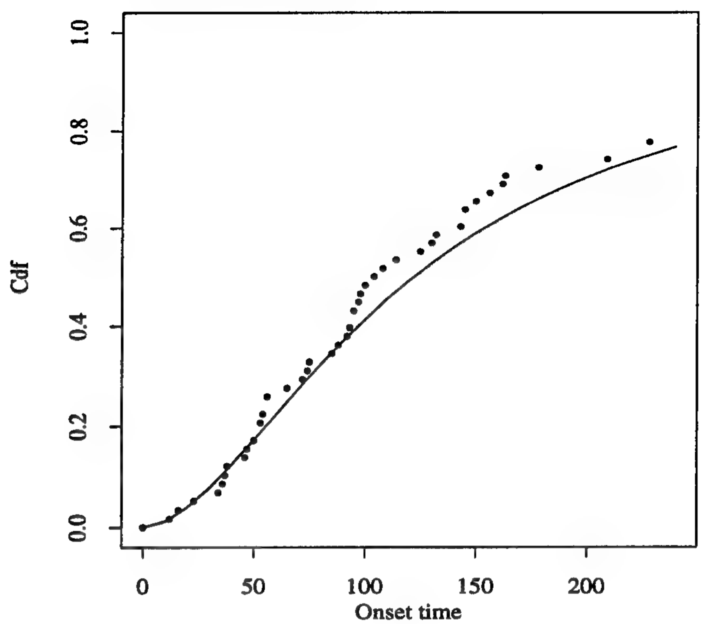
Clearly, the predicted models in all the above examples agree closely with the empirical probabilities from the DCS database. We have observed that this model with the different strata does perform better than the earlier model. The earlier



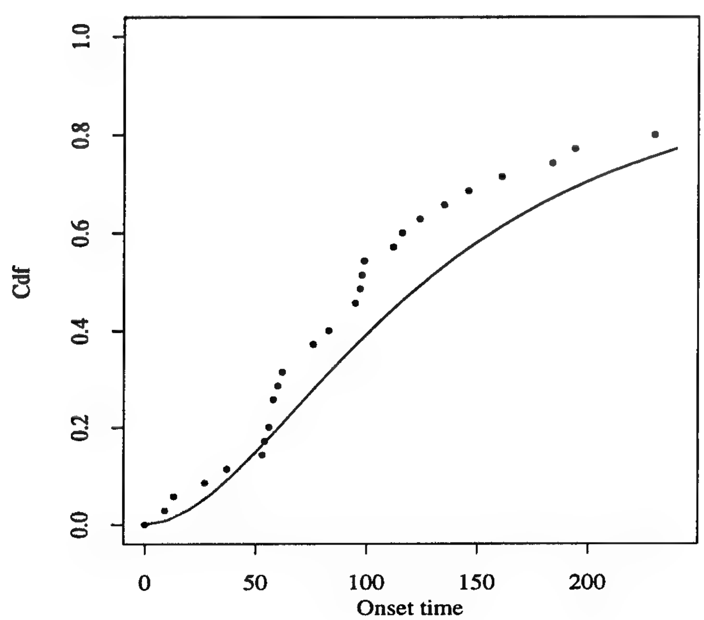
(a) PRES=179 mm Hg, BR=75 min, TALT= 180 min, EX



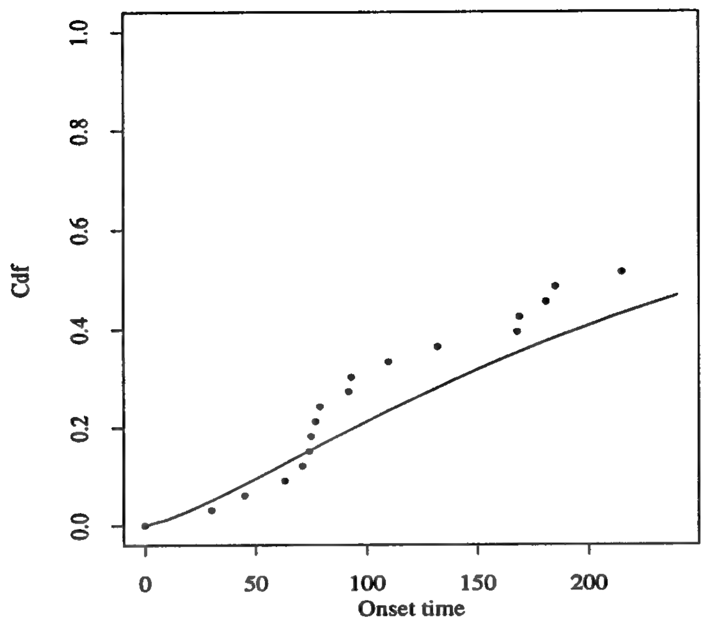
(b) PRES=179 mm Hg, BR=75 min, TALT= 180 min, REST



(c) PRES=226 mm Hg, BR=60 min, TALT= 240 min, EX



(d) PRES=282 mm Hg, BR=0 min, TALT= 180 min, EX



(e) PRES=314 mm Hg, BR=0 min, TALT= 240+ min, EX

model was not able to predict adequately for zero prebreathe cases. Figure (d) for the 25,000 ft profile clearly indicates a good fit. This is not surprising in the light of the remarks about threshold and interaction effects discussed earlier.

Exercise during Preoxygenation

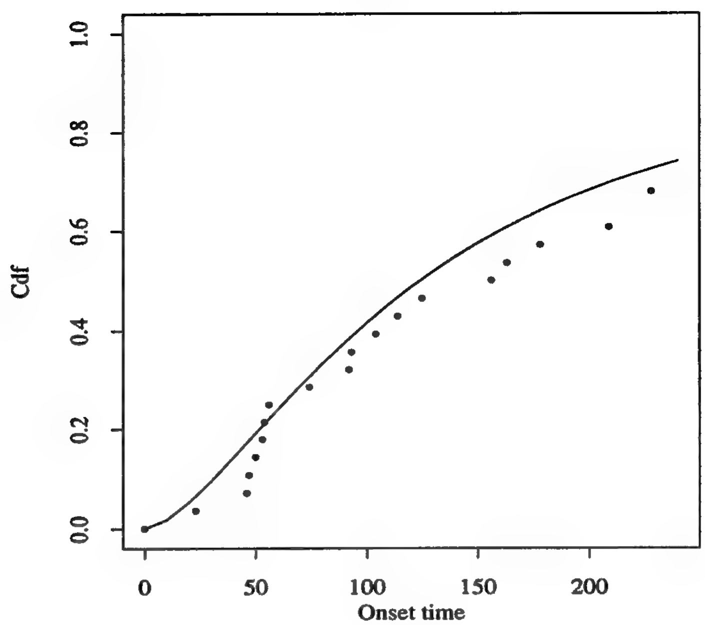
Preoxygenation using 100 % oxygen is a common method used to remove nitrogen from the tissues, and hence reduce the risk of DCS. Based on the SAS outputs given above, it is evident that preoxygenation time is a significant risk factor, especially at moderate and high altitudes. The longer the prebreathe duration, the lower the risks of DCS. Comparisons of profiles with 0 and 60 minutes of prebreathe show significantly reduced rates of DCS for the group that prebreathed for 60 minutes However for 90 minutes of prebreathe, the reduction over the 60 minute duration is not that significant. This suggests, perhaps a law of diminishing returns with

increased preoxygenation times.

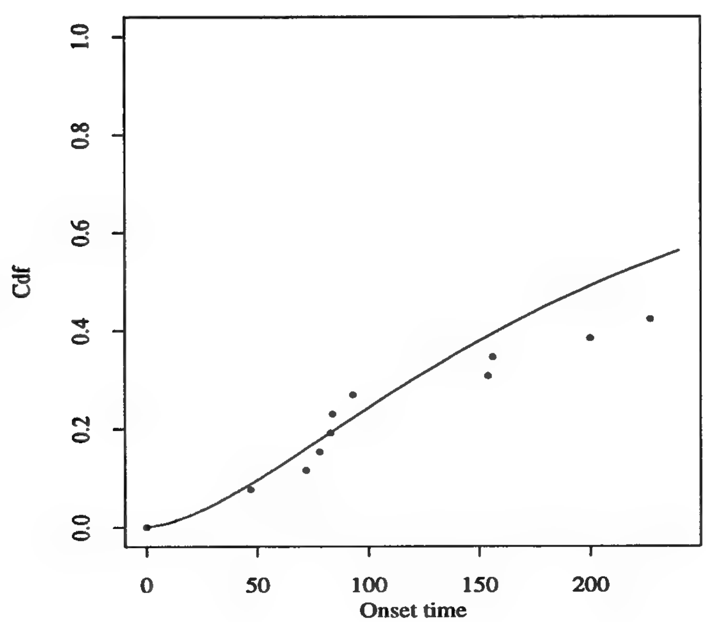
It has been hypothesized that exercise during the preoxygenation, can increase the efficiency of denitrogenation. To test this hypothesis, chamber experiments have been carried out using 10 and 15 minutes of exercise during the initial part of the prebreathe. Webb et al [17] have shown that incidence of DCS following 1-hour of prebreathe with 10 minutes of exercise was significantly lower than the incidence of DCS for the 1-hour resting profile. The incidence of DCS for 15 minutes of exercise was, however, not significantly different from the 1-hour resting profile. This may suggest that longer exercise durations are not necessarily effective in denitrogenation, and in fact may hasten the onset of symptoms.

To evaluate the effects of exercise during prebreathe, we added another risk factor in the model. This new variable, BREX is defined as $exp(-TIME/BR)$, where $TIME$ is the number of minutes of exercise. For profiles with no exercise during the prebreathe, this variable was set to 0. Since this experiment was conducted only at 30,000 ft, only that data was used to fit the loglogistic model. Using the model, we predicted the probability of DCS for the resting and prebreathe-exercise profiles. The graphs are shown below. Clearly, the inclusion of this factor explains the difference in DCS rates for the two protocols. The graphs also show that the overall incidence of DCS is reduced. Another interesting observation is that prebreathe with exercise seems to result in delayed DCS onset times for individuals who do bend. The most obvious advantage of exercise during prebreathe is the fact that it reduces preoxygenation duration while offering the same protection against DCS. Operationally, this is a very viable option for high altitude flights and EVA applications.

Since only data at 30,000 feet is available at present, further analysis would need to be conducted to accurately measure the effects of exercise. As more data becomes available, it would eventually be of interest to include this risk factor into any predictive model.



(f) Resting Prebreathe, 60 min



(g) Prebreathe with exercise, 10/50 min

Conclusions

In this report, we have considered models for predicting the probability of DCS symptoms over time. The loglogistic distribution seems most appropriate in describing the phenomenon of altitude DCS. The model predicts the probability/ risk of DCS as a function of several risk factors, including pressure, preoxygenation time, and exercise. Since the effects of exercise vary with altitude, it was decided to create separate models for different ranges of pressure. The predictions from these models agree closely with empirical data. Models that include the effects of exercise during prebreathe have also been considered. In addition, a preliminary analysis of the effects of inflight denitrogenation have also been attempted.

Even with over 2000 records in the database, there are combinations of pressure, exercise, and preoxygenation times with little or no data. Predicting DCS probabilities for these profiles clearly becomes a challenge. As more experimental data becomes available, the model will need to be constantly evaluated and modified. The Air Force Research Laboratory at Brooks AFB is currently involved in a validation study. Data from these experiments will be used to provide an assessment of the model and valuable information on gaps in the database. The data from the validation study should be available by next summer, and will provide the basis for a final mathematical model capable of predicting DCS over a wide spectrum of operational profiles.

Acknowledgments:

The author would like to thank Dr. Andrew A. Pilmanis, Dr. James T. Webb, Dr. Kevin M. Krause, and Mr. Lambros Petropoulus for their assistance. The author acknowledges the many helpful suggestions and comments received during this project.

Bibliography

P. Weathersby, L. Homer, and E. Flynn, "On the likelihood of decompression sickness," *Journal of Applied Physiology*, vol. 57, pp. 815–825, 1984.

P. K. Weathersby, S. Survanshi, L. Homer, E. Parker, and E. Thalmann, "Predicting the time of occurrence of decompression sickness," *Journal of Applied Physiology*, vol. 72, pp. 1541–1548, 1992.

H. Van Liew, J. Conkin, and M. Burkard, "Probabilistic model of altitude decompression sickness based on mechanistic premises," *Journal of Applied Physiology*, vol. 76, pp. 2726–2734, 1994.

W. Gerth and R. Vann, "Statistical bubble dynamics algorithms for assessment of altitude decompression sickness incidence," tech. rep., Armstrong Laboratory, 1995.

K. Kumar, J. Waligora, and D. Calkins, "Threshold altitude resulting in decompression sickness," *Aviation, Space, and Environmental Medicine*, vol. 61, pp. 685–689, 1990.

K. Kumar, D. Calkins, J. Waligora, J. Gilbert III, and M. Powell, "Time to detection of circulating microbubbles as a risk factor for symptoms of altitude decompression sickness," *Aviation, Space, and Environmental Medicine*, vol. 63, pp. 961–964, 1992.

K. Kumar, J. Waligora, J. Gilbert III, and M. Powell, "Epidemiology of decompression sickness under simulated space extravehicular activities," *Aviation, Space, and Environmental Medicine*, vol. 64, pp. 1032–1039, 1993.

K. Kumar and M. Powell, "Survivorship models for estimating the risk of decompression sickness," *Aviation, Space, and Environmental Medicine*, vol. 65, pp. 661–665, 1994.

J. Conkin, K. Kumar, M. Powell, P. Foster, and J. Waligora, “A probabilistic model of hypobaric decompression sickness based on 66 chamber tests,” *Aviation, Space, and Environmental Medicine*, vol. 67, pp. 001–008, 1996.

N. Kannan, Raychaudhuri, and A. Pilmanis, “A loglogistic model for altitude decompression sickness,” *Aviation, Space, and Environmental Medicine*, vol. 69, pp. 965–970, 1998.

N. Kannan, “Predicting altitude decompression sickness using survival models,” tech. rep., Air Force Office of Scientific Research, 1997.

J. Lawless, *Survival Models and Methods for Lifetime Data.* Wiley, New York, 1982.

E. Lee, *Statistical Methods for Survival Data Analysis.* Wiley, New York, 1992.

J. Webb, R. Krutz, and G. Dixon, “An annotated bibliography of hypobaric decompression sickness research conducted at the crew technology division, usaf school of aerospace medicine, brooks afb, texas from 1983-1988,” Tech. Rep. USAFSAM-TP-88-10R, USAF School of Aerospace Medicine, 1990.

Z. Sulaiman, A. Pilmanis, and R. O’Connor, “Relationship between age and susceptibility to decompression sickness,” *Aviation, Space, and Environmental Medicine*, vol. 68, pp. 695–8, 1997.

M. Ryles and A. Pilmanis, “The initial signs and symptoms of altitude decompression sickness,” *Aviation, Space, and Environmental Medicine*, vol. 67, pp. 983–989, 1996.

J. Webb, M. Fischer, C. Heaps, and A. Pilmanis, “Exercise-enhanced preoxygenation increases protection from decompression sickness,” *Aviation, Space and Environmental Medicine*, vol. 67, pp. 618–624, 1996.

AIRCRAFT AND DMT: MODELING AND ANALYSIS OF TRAINING EFFECTIVENESS, FLIGHT TRADEOFFS, COSTS AND RESOURCE ALLOCATIONS

Ram Ramesh
Associate Professor
Department of Management Science & Systems

School of Management
State University of New York at Buffalo
Buffalo, NY 14260

Final Report for:
Summer Faculty Research Program
Airforce Research Laboratory, Mesa, AZ

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force base, DC

And

Air Force Research Laboratory
Aircrew Training Division
Mesa, AZ

September, 1998

Aircraft and DMT: Modeling and Analysis of Training Effectiveness, Flight Tradeoffs, Costs and Resource Allocations

Ram Ramesh
Department of Management Science & Systems
School of Management, SUNY at Buffalo

Abstract

Distributed Mission Training (DMT) is a revolutionary training paradigm currently evolving at the Department of Defense, especially at the Air force. The fundamental technologies on which DMT is built are: virtual simulation, networked war gaming and intranets. The rationale for DMT is derived from the characteristics of contemporary warfare and the increasing emphasis on creating technology based training environments that realistically capture the complexities and demands of modern military operations. While the dimensions and complexity of modern warfare are expanding, the ability of the defense services to train forces in a realistic environment is being increasingly constrained. The primary constraints arise from limited resources for team skill training using actual equipment such as aircraft, safety limitations of live training events such as air-to-air missiles for instance, and security constraints due to operational conditions. Consequently, DMT is strongly emerging as an alternate but effective mode of team training in the defense services. In this research, we develop models and a spreadsheet decision support system to address the following key questions concerning DMT: (i) *Should DMT be deployed at all, as part of continuation/replacement training programs for F16 crew, (ii) What is the extent to which continuation/replacement training can be conducted using DMT systems, (iii) What are the various system configurations under which DMT can be deployed, (iv) What are the specific costs associated with DMT systems, and (v) Are there specific measures of effectiveness that can they be used to evaluate the potential DMT configurations up front.*

Aircraft and DMT: Modeling and Analysis of Training Effectiveness, Flight Tradeoffs, Costs and Resource Allocations

Ram Ramesh
Department of Management Science & Systems
School of Management, SUNY at Buffalo

1. Introduction

Distributed Mission Training (DMT) is a revolutionary training paradigm currently evolving at the Department of Defense, especially at the Airforce. The fundamental technologies on which DMT is built are: *virtual simulation*, *networked war gaming* and *intranets*. While the United States is pioneering at the cutting edge of these technologies, several other countries such as U.K., France, Israel and the Netherlands are also seriously engaged in such initiatives. DMT is emerging as the technologies of the next generation training systems. This can be observed from (1) the keynote address by General Richard E. Hawley, Commander-in-Chief, Air Combat Command of the US Airforce at the 19th Industry/Interservice conference (1998), (2) the proceedings of the Industry/Interservice conferences, 1997, 1998 and (3) the following web pages: http://www.tno.nl:8080/instit/fel/div2/feldiv23.html,
http://stbbs.wpafb.af.mil/STBBS/naecon/naeconhtm/al/nan11-12/sld001.htm,
http://huachuca-usaic.army.mil/SCHOOL/DOTD/dotdhpold.html.

The rationale for DMT is derived from the characteristics of contemporary warfare and the increasing emphasis on creating technology based training environments that realistically capture the complexities and demands of modern military operations. For example, the range and performance of modern aerial weapons systems enable domination of larger areas. Tighter linkages between sensors like AWACS or JSTARS and shooters like fighters and bombers require increased emphasis on teamwork for successful mission execution. Tactics are increasingly based on the technology and behavior of an adversary's integrated defense systems rather than individual platforms. Taken together, these trends are blurring the distinction between operational and tactical levels of training and mission preparedness. DMT is an evolution from this rapidly advancing technology driven warfare, and is based on the following key observations:

- Teams, not individuals execute missions. For example, a flight of four F-15s working together with other teams - an AWACS and several flights of F-16 strike aircraft - perform an offensive counter air mission.

- Team skills are built upon, but different from, individual skills. For example, proficiency in basic flight maneuvers, an individual skill, is necessary but not sufficient for proficiency in air combat maneuvering, which is a team skill.

While the dimensions and complexity of modern warfare are expanding, the ability of the defense services to train forces in a realistic environment is being increasingly constrained. The primary constraints arise from limited resources for team skill training using actual equipment such as aircraft, safety limitations of live training events such as air-to-air missiles for instance, and security constraints due to operational conditions. Consequently, DMT is strongly emerging as an alternate but effective mode of team training in the defense services. Distributed networks of advanced simulators of various military equipment such as fighters, bombers, battle tanks and even ships and aircraft carriers are being developed with a principal focus on team training. While the airforce spearheads this effort, linkages of such a variety of simulators are steadily developing, leading to novel and effective joint forces team skill development exercises.

DMT is still very much the state-of-the-art. It is expected to become the state-of-practice in the next millennium. Currently, distributed networks linking four ship F-15 simulators at each DMT node are in different stages of development in the airforce. They will be deployed, tested and analyzed for training effectiveness at Langley, Eglin and Tinker airforce bases. They will also be linked with AWACS simulators at the network nodes to study the effectiveness of joint training programs. The current DMT trainer web under development is a secure DoD intranet hosting virtual and constructive simulations for F-15, F-16, and other logistics support. As more weapons systems join the environment, DMT will be able to support a joint synthetic battlespace accommodating many virtual, live and constructive simulations, leading to theater-specific interservice training exercises.

2. DMT Research Issues

The DMT research issues underlying this study are best stated from the point of view of COMMACC, who ultimately has to make the decisions concerning the deployment of the simulator systems as part of field training. These decisions can be best made by first addressing the following strategic and operational level questions regarding simulator deployment.

1. ***Should DMT be deployed at all, as part of continuation/replacement training programs for F16/F15 crew at ACC?***

The answer to this question is just a simple YES or NO. If the answer is NO, then it probably implies a continuation of existing flight training systems that are currently in use. However, in order to answer this question, we need to investigate the following questions concerning the DMT system.

1. ***What is the extent to which continuation/replacement training can be conducted using DMT systems?***

This requires a categorization of the training tasks into:

- Tasks that DO NOT require training on DMT. This breaks into:
    - Tasks that are regularly practiced in the aircraft sorties
    - Tasks that SHOULD be learnt only in aircraft sorties
- Tasks that can POSSIBLY be trained in DMT. These tasks can be categorized on a 0-1 scale, mapping the range between "No Possibility" and "Full Possibility " of training. This range identifies (in some discrete measure), tasks that can be entirely transferred to DMT, tasks that can be partly transferred and tasks that can not be transferred with any significant training effectiveness.
- Tasks that should be trained ONLY on simulator devices.

A good starting point for this analysis is the RAP mission tasklist. The above categorization will reveal the training possibilities with the simulators. However, in order to assess the training effectiveness with DMT, specific studies either with subject matter experts (SME) or live field experimentation with DMT are needed.

2. ***What are the various system configurations under which DMT can be deployed?***

This can be addressed from two broad perspectives: a *strategist* or a *designer*. While a designer needs a comprehensive analysis of the numerous system architectures to select one that is most appropriate, a strategist would be interested in a few, broadly defined configurations. Full architectural specifications would take the form of detailed matrices describing the various functions and features of the system and the designer's options in selecting them (see (1)). Putting the functions and features together and organizing into some sort of classification such as "Essential", "Preferable", and "Optional" with an assessment of their costs and benefits would give the designer a full range of choices in optimally configuring a simulator system. However, a strategist may not need such an abundance of detail, but would like to explore a few, significant, high-level design alternatives to choose from.

3. ***What is the extent to which continuation/replacement training be conducted in each configuration of DMT?***

This question revisits questions 2 and 3 and explores the relationships between simulator configurations and transfer of training tasks. A complete mapping of the broadly defined DMT configurations (developed for question 3) and the tasks that can be transferred to the simulators (developed for question 2) would show for each simulator configuration:

- What tasks are transferable
- What is the estimated transfer effectiveness

Models of overall training effectiveness under each simulator configuration (maybe, obtained from a template model for all configurations) need to be developed to answer this question.

4. ***How should the training curricula be modified (improved?) to accommodate the simulator training component in the overall existing training program?***

Clearly, the induction of DMT will change the ways in which training is conducted. Therefore, for each potential configuration, at least an approximate training curricula needs to be developed. However informal or ill-structured the training may be, it is necessary to organize the training tasks in some high-level framework to determine the effectiveness of simulator training. Consequently, it will be possible to evaluate each configuration jointly with its associated training curricula for an overall analysis of its effectiveness. This question addresses the effects of introducing simulators in a training program in terms of curricular changes, new tasks that can be trained etc.

5. ***How should the DMT systems be located and used among the various training units of ACC? What are the operational and performance implications of this?***

This is a location and usage modeling issue, to be addressed in conjunction with the configurations and curricular issues. Note that there are several ways in which the simulators can be configured and used. This question addresses the key decisions to be made in introducing simulator training in the ACC units. These decisions are: *DMT configurations, location of the DMT systems* and *the extent of their usage*. In turn, these decisions will affect: the training curriculum (incorporating simulator training), training schedules at the units, throughput of trained personnel (a measure of quantity) and the levels of skills acquired (a measure of quality). Therefore, it is important to study the training systems currently in use, the training curricula, simulator configurations and their capabilities, wing/squadron level training event scheduling processes, measures and systems for performance evaluation and throughput assessment.

6. ***What are the operational and maintenance requirements for the simulators?***

These requirements can be specified in terms of: *system support, manpower support, spare parts/maintenance support* and *organizational/funding support.*

7. ***What are the specific costs associated with the simulators? How does each potential configuration evaluate in this dimension?***

Using the Knapp & Orlansky model, we can organize all the costs associated with simulator systems into three categories: *Research & Development Costs, Initial Investment Costs and Operating* and *Support Costs.* While the K&O model is rather comprehensive, we will need to fine tune it to those cost elements that are relevant and measurable/estimable in the present context. See the C130 study (2) for a baseline modeling approach for this. In a cost evaluation framework, it is necessary to build a basic template model, and assess each configuration separately from this.

8. ***Are there specific measures of effectiveness? If so, how can they be used to: (i) evaluate the potential configurations up front, and (ii) maintain a continuous test and evaluation of the configuration chosen for implementation during its life cycle?***

At the outset, two important measures that appear prominently are:

- Release of sortie time due to the transfer of training to simulators
- Additional skills that can be acquired from simulator training. We assume that these skills are not currently acquired due to the nonavailability of appropriate DMT level training. This measure clearly points to new training possibilities arising out of the DMT systems.

9. ***How does the policy of introducing simulator training relate to other USAF policies (both current and evolving) regarding training in general, training systems, and future directions? Is the DMT system sufficiently sustainable over a significant time horizon over which several policy changes, new initiatives and directions could evolve? What is the projection of these future events? What are the likelihoods of at least some of the major projections?***

This question deals with long range planning issues. It is very important to recognize their impact on the proposed training systems. A model projecting the future and life cycle of the DMT systems is critical for this purpose.

10. Based upon the above analyses, if we decide upon: (i) an optimal DMT configuration, (ii) a modified training plan incorporating training with the chosen configuration, and (iii) a location and usage plan for the simulators, Then:

- ***What is our long range plan in terms of tasks to be carried out, schedules and funding requirements?***
- ***What are our short range plans leading to our long range goals?***
- ***How can we ensure that the near term activities roll into the long range plan?***

3. Current Research Scope and Objectives

The above issues lay out a broad framework research on DMT systems, DMT based training program development and the field deployment of DMT. While this framework addresses the major strategic decisions in DMT development and application, a full scale analysis of all the strategic issues and the development of solutions to their underlying problems is beyond our current scope. Therefore, we restrict our research to the following areas in order to develop a proof-of-concept decision support system that can be subsequently extended over the full range of decisions.

- Continuation/replacement training in F16 airframes.
- Assessment of F16 training tasks in terms of their trainability on (i) aircraft only, (ii) DMT only, and (iii) aircraft as well as DMT.
- High level assessment of the critical components of DMT under relevant and important configurations. Broadly, we consider 2-ship and 4-ship DMT systems in this analysis.
- Assessment of the trades from aircraft flying time to DMT flying time that yield the same level of combat mission readiness in tasks that are trainable in both the media.
- Assessment of the training capacities yielded by a set of aircraft and DMT training resources under various strategies jointly utilizing aircraft and DMT together for training.
- High level assessments of the costs associated with the DMT configurations.
- Development of a spreadsheet based parametric sensitivity analysis model to perform (i) aircraft - DMT flying time tradeoff analysis, (ii) Training capacity analysis for joint aircraft - DMT training, and (iii) high level cost analysis of DMT configurations. This model puts all the above analyses together in the framework of a decision support system.

Putting it all together, this research addresses the questions (1) - (4), (8) - (9) above for F16 training at a high level. Although the analyses carried out in this regard are specific to F16 training,

they can be generalized to training in any airframe. The F16 airframe is only used to demonstrate the analyses and the systems developed, and serves as the proof-of-concept for the generalizable ideas developed in this work. The questions (5) - (7) specifically pertain to training program design and the operational logistics. The questions (10) - (11) pertain to high level strategic decision making in training doctrines development, training program design and the acquisition of training systems and services. Although these questions are critically important, they are beyond the scope of the current work, and should be studied in future research.

4. Research Plan and Methodology

The key research resources utilized in this study are: (i) the DMT expertise available at the Airforce Research Laboratory (ARL), Mesa, AZ, ACC/DO and ACC/DR at Langley Airforce base, and ASC/YW at the Wright Patterson Airforce base, (ii) the four ship DMT prototypical system currently at ARL, Mesa, and (iii) documented sources of information on DMT, distributed training environments at the Airforce, and prior cost benefit studies on multiship training on different airframes (see Moor and Andrews(3, 4), Mudd et al. (6) and Derrick and Davis (6) among others). The DMT systems are in early stages of their life cycle development, and none have reached the level of field deployment yet. Since most of the information needed in this research occurred in informal domains such as human expertise for instance, we had to rely on the assessments of the numerous subject matter specialists who participated in this study. Accordingly, our research methodology consisted of the following phase of study:

- ***Phase 1:*** Initial data collection on DMT characteristics, training requirements and strategies, and cost/performance indicators. This has been accomplished through a series of interviews with experts at ARL, ACC/DO, ACC/DR and ASC/YW. This phase resulted in a broad framework of the subsequent data collection on the structural details of DMT based training.
- ***Phase 2:*** Refinement of the data collection framework. This involved a series of interviews with the experts at ARL and a detailed study of the literature on cost-benefit studies on multiship as well as single ship simulator training (3,4,5,6). Putting all these studies together, a systematic and controlled data collection strategy using subject matter experts and the DMT prototype at ARL emerged.
- ***Phase 3:*** The Roadrunner'98 exercises and data collection. ARL conducted these exercises between July 13 - 20, 1998 at Mesa. Several F16 pilots were brought to Mesa to conduct joint war gaming exercises with AWACS and F15 air superiority support using distributed simulation. The four ship F16 DMT at Mesa has been networked with other remote simulation sites, and the war games were conducted as script-controlled scenarios with both constructive and live simulation events in joint exercises. The primary objective of Roadrunner'98 has been

to explore the training possibilities with DMT, its training effectiveness as compared to aircraft training, and highlight its strengths and weaknesses that would guide future R&D in this area and its induction into the Airforce training doctrine. Our study has been part of these exercises. Six questionnaires were designed for our study and administered to 15 subject matter experts after they have had significant first hand experience with the DMT systems. The assessments they provided forms a substantial basis for the training effectiveness analyses embedded in te decision support system developed in the next phase.

- ***Phase 4:*** The development of a spreadsheet based decision support system to evaluate the costs and benefits of DMT systems and perform parametrical sensitivity analyses. This system has been developed using MS Excel, and demonstrated to the ARL personnel.

We develop the models underlying the decision support system and its architecture in the following section.

5. The Decision Support System

We present the following details on the elements of the DSS in this section: key elements of the data collected, assumptions underlying the analyses, broad structural details of the models embedded in the system and a summary of the system outputs.

5.1. System Database

The database consists of two components: training effectiveness data and cost data. The effectiveness data has been collected from the six questionnaires used in the Roadrunner study. The questionnaires employ a set of 45 training events in F16 combat training. These events, referred to as *Mission Elements*, have been identified using the F16 RAP and the collective knowledge of the ARL personnel. The key elements of this database are:

- A magnitude scale rating of the mission elements on their importance to combat readiness (Questionnaire 1)
- An evaluation of the training experience with the aircraft to actual combat experience using a 0-4 scale (Questionnaire 2)
- An evaluation of the training experience with 2-ship DMT to actual combat experience using a 0-4 scale (Questionnaire 3)
- An evaluation of the training experience with 4-ship DMT to actual combat experience using a 0-4 scale (Questionnaire 4)

- An evaluation of the minimum number of sorties required for combat readiness of inexperienced pilots when: (i) only the aircraft is used, (ii) aircraft and 2-ship DMT are used, and (iii) aircraft and 4-ship DMT are used. The experts specified the break up between the aircraft and DMT sorties in (ii) and (iii). This questionnaire obtained the tradeoff data between aircraft and DMT sorties in the case of inexperienced pilot training (Questionnaire 5)
- An evaluation of the minimum number of sorties required for combat readiness of experienced pilots when: (i) only the aircraft is used, (ii) aircraft and 2-ship DMT are used, and (iii) aircraft and 4-ship DMT are used. The experts specified the break up between the aircraft and DMT sorties in (ii) and (iii). This questionnaire obtained the tradeoff data between aircraft and DMT sorties in the case of experienced pilot training (Questionnaire 6)

The cost database has been compiled from detailed discussions with the ARL personnel and other relevant data sources available at the laboratory. The key elements of this database are:

- Nonrecurring initial fixed costs
- Long term recurring fixed costs
- Direct operating costs
- Indirect operating costs

Putting training effectiveness and costs together, the models embedded in the system provide a direct cost - benefit analysis of 2-ship and 4-ship DMT systems under various levels of DMT usage.

5.2. Models of Training Effectiveness Analysis

The parameters of training effectiveness analysis that are derived from the Roadrunner questionnaires are as shown in Table 1.

5.2.1. Modeling Objectives

The objectives in the development of the models of training effectiveness are as follows.

- Determine the training effectiveness of the tasks that can be trained on the aircraft, 2-ship DMT and 4-ship DMT, respectively {EFAC, EF2D, EF4D}.
- Determine the minimum practice requirements on the mission elements for combat readiness when: (1) No DMT is available, (2) 2-ship DMT is available, and (3) 4-ship DMT is available. Determine these requirements for both inexperienced and experienced pilots {INAC, IN2D_AC, IN2D_2D, IN4D_AC, IN4D_4D, EXAC, EX2D_AC, EX2D_2D, EX4D_AC, EX4D_4D}.
- Model the interrelationships among these parameters and validate: (1) consistency among the parametric relationships, and (2) consistency among the evaluating subject matter experts.

- Model overall training effectiveness when: (1) No DMT is available, (2) 2-ship DMT is available, (3) 4-ship DMT is available. Develop measures for:
    - Composite overall training effectiveness for each design of the training systems configuration (No DMT, with 2-ship, with 4-ship)
    - Training effectiveness accruing from each system (aircraft, 2-ship, 4-ship) in each design configuration
    - Differential training effectiveness of each system on tasks that are uniquely trainable only on the individual system concerned. We assume that a 4-ship is a superset of a 2-ship, and hence, there is no differential of 2-ship over 4-ship (that is, all tasks that can be trained on a 2-ship can also be trained on a 4-ship)
        - Aircraft over 2-ship and 2-ship over aircraft
        - Aircraft over 4-ship and 4-ship over aircraft
        - 4-ship over 2-ship

Table 1. Parameters of Training Effectiveness Analysis

1. Task Importance: IMP(k), k=1,...,45 <Magnitude scale assessment>
2. Aircraft Trainability/Effectiveness: EFAC(k), k=1,...,45 <scale: 0 - 4>
3. 2-Ship DMT Trainability/Effectiveness: EF2D(k), k=1,...,45 <scale: 0 - 4>
4. 4-Ship DMT Trainability/Effectiveness: EF4D(k), k=1,...,45 <scale: 0 - 4>
5. Comparative Assessment - Inexperienced Pilots:
5.1. No DMT Available:
Aircraft Sorties: INAC(k), k=1,...,45
5.2. 2-Ship DMT Available:
Aircraft Sorties: IN2D_AC(k), k=1,...,45
2-Ship DMT Sorties: IN2D_2D(k), k=1,...,45
5.3. 4-Ship DMT Available:
Aircraft Sorties: IN4D_AC(k), k=1,...,45
4-Ship DMT Sorties: IN4D_4D(k), k=1,...,45
6. Comparative Assessment - Experienced Pilots:
6.1. No DMT Available:
Aircraft Sorties: EXAC(k), k=1,...,45
6.2. 2-Ship DMT Available:
Aircraft Sorties: EX2D_AC(k), k=1,...,45
2-Ship DMT Sorties: EX2D_2D(k), k=1,...,45
6.3. 4-Ship DMT Available:
Aircraft Sorties: EX4D_AC(k), k=1,...,45
4-Ship DMT Sorties: EX4D_4D(k), k=1,...,45

- Model the transfer functions from the aircraft to 2-ship and 4-ship DMTs, and derive the following estimations:

- Estimation of the transfer functions for the mission elements
- Estimation of the composite aircraft-DMT transfer curve
- Estimation of the trainee capacities in each aircraft-DMT joint training configuration
- Estimation of the level of training that can be accomplished with a training configuration, given resource constraints and training loads

- Develop spreadsheet based decision models for selecting an appropriate training systems configuration, their deployment and usage for a given set of training requirements and loads at a wing.

5.2.2. Modeling Training Effectiveness

We first normalize all SME input data. The normalized measures are then averaged over all the SMEs and used in the subsequent models.

5.2.2.1. Transfer of Training Estimation by Mission Elements

Consider a mission element k. Without loss of generality, we will use the following generalized notations in describing the transfer of training models:

AC(k) = # of aircraft sorties needed if no DMT is available.

D_AC(k) = # of aircraft sorties needed if a DMT is also used

D_D(k) = # of DMT sorties needed.

In this notation, we have suppressed (1) the types of DMT and (2) types of training (inexperienced/experienced) in the above notation for simplicity. The four categories of training configurations (IN/2SHIP, IN/4SHIP, EX/2SHIP, EX/4SHIP) by replacing these generic parameters with their respective parameters in the following models.

We model the transfer of training using two dimensions: # of aircraft sorties and # of Sim sorties. We have two points on this transfer curve from the SME data as follows:

(0, AC(k)) and (D_D(k), D_AC(k)). We denote the point (0, AC(k)) as the case where *<u>no DMT is used</u>*, and the point (D_D(k), D_AC(k)) as *<u>the limiting case of DMT use as specified by the SMEs</u>*. We assume the commonly used exponential transfer function (see Bickley []). The exponential function has been very well studied in the literature, and has been in wide use in the training area. The transfer function in this case is modeled as $y = Ae^{-Bx}+C$, where

x = # of Sim sorties

y = # of aircraft sorties

A, B, C = transfer function constants

Using the two points along this curve available from the SME data and the transfer effectiveness of aircraft (EFAC(k)) determined from the SMEs earlier, we determine A, B and C as follows.

A_k = {EFAC(k)}{AC(k)}

C_k = {1-EFAC(k)}{AC(k)}

Now, plugging in the other SME point {x = D_D(k), y = D_AC(k)}on the transfer curve, we get B_k as:

B_k = -{1/D_D(k)}{ln[{1/EFAC(k)}{(D_AC(k)/AC(k))+EFAC(k)-1}]}

For the sake of simplicity, we will denote the transfer curve for mission element k as:

$\mathbf{y}_k = A_k e^{-B_k x_k} + C_k.$

5.2.2.2. Composite {Aircraft, DMT} Transfer of Training Estimation

We now turn our attention to the determination of an *overall transfer curve*: from total number of aircraft sorties to total number of DMT sorties, putting all the missions together. Clearly, this is a very complex issue, as (1) many mission elements could be performed in a mission sortie, and (2) a mission element could be needed in many missions. However, from the point of view of *estimation*, we assume that the number of sorties indicated by the SMEs in each category represent the approximate proportion of the time a pilot is required to spend in each mission element during training for combat ready preparation. Consequently, we deal with the normalized percentage values of the sorties requirements in the following analysis.

Consider a two dimensional plot of normalized total aircraft sorties time versus normalized total DMT time. We approximate sorties data for time, as the analysis is considered for a long run period such as a year. Consider any training system configuration with x(k) and y(k) sorties used for mission element k on the DMT and aircraft, respectively. These two parameters are the same as those defined in the transfer curve estimation above. Let TY denote the ratio of the total time actually spent in aircraft training when DMT is used to the total time when no DMT is used. Similarly, let TX denote the ratio of the total time actually spent in DMT training when DMT is used to the total time when DMT is used in the limiting case as specified by the SMEs. Using this, we define TY and TX as follows:

TY = $\Sigma_{\kappa=1,45}$ y(k)/$\Sigma_{\kappa=1,45}$ AC(k)

TX = $\Sigma_{\kappa=1,45}$ x(k)/$\Sigma_{\kappa=1,45}$ D_D(k)

When TX = 0, the value of TY = 1. This corresponds to the case where no DMT is used. Similarly, when TX = 1, TY = $\Sigma_{\kappa=1,45}$ D_AC(k)/$\Sigma_{\kappa=1,45}$ AC(k). This corresponds to the limiting case of DMT use as specified by the SMEs. Hence, TX and TY range between 0 and 1 in this normalized plot.

For any value of TX between 0 and 1, we can determine the corresponding total aircraft time required (TY) from the individual transfer functions developed in the above analysis. However, we can get into a serious combinatorial problem leading to inconsistencies in estimating the total

times when they are *assembled* from individual mission element transfer functions. Hence, we suggest the following procedure to systematically capture the tradeoffs.

First, starting from (TX=0, TY=1), consider transfers from aircraft to sim in steps of P%. For instance, the first point on this curve is aggregated from a decrease in aircraft time in **all** the mission elements by P% of the difference between the maximum and minimum aircraft practice requirements specified by the SMEs {AC(k)-D_AC(k)}. Continue this as far as possible. Finally, join all the points thus generated with a smooth curve. We call this transfer curve as a ***P% step reduction curve,*** as this stepwise reduction is universally applied to all the mission elements. Clearly, there are numerous other ways to effect these transfers, and each transfer would produce a curve. Such combinations can best be analyzed using a spreadsheet. The P% points on the normalized plot are determined by the following procedure.

1. Select P% (say around 20%). Set n =1.
2. y(k, nP%) = AC(k) - (nP){AC(k)-D_AC(k)}, k = 1,...,45
3. x(k, nP%) = {-1/B_k}ln{(y(k,nP%) - C_k)/A_k}, k=1,...,45
4. Calculate TX(nP%) and TY(nP%) from the above definitions.
5. If y(k, nP%) $\leq$ D_AC(k) stop. Else go to step 6.
6. Set n = n+1. Return to step 2.

5.2.2.3. Aircraft/DMT Trainee Capacity Estimation

Recall that (1) many mission elements could be performed in a mission sortie, and (2) a mission element could be needed in many missions. Therefore, the estimation of the actual sorties requirement for a given trainee load and training requirements from the data we have is rather difficult. However, we can estimate lower bounds on the number of pilots that can be trained using the existing aircraft and DMT resources over a period of time, say a year. Then using a *mission element bundling* concept, we can derive a sensitivity analysis on the estimated sorties requirements. This analysis is developed as follows.

Let N_AC and N_D denote the number of aircraft and DMT sorties available in a year for training. To begin with, we make the following assumptions:

1. All aircraft sorties are flown as 4-ship, in order to establish a common basis for our comparative analysis.
2. The number of times the mission elements are performed in the aircraft and DMT will be distributed according to the proportions indicated by the SME data.
3. If the total number of sorties required for combat readiness (as specified by the SMEs) are not available (either in the aircraft or DMT), then whatever is available will be used, and the sorties will be distributed among the missions along the proportionality assumption above.

4. The set of sortie requirements on the mission elements can be *bundled* into a set of sortie requirements on missions.

Now, based on assumptions 3 and 4, we introduce two parameters as follows:

- An **aircraft sortie reduction factor** δ_a, which ranges between 0 and 1, indicating the level to which the required total number of aircraft sorties for a given training load that can be accomplished with the available number of aircraft sorties.
- A **DMT sortie reduction factor** δ_s, which ranges between 0 and 1, indicating the level to which the required total number of DMT sorties for a given training load that can be accomplished with the available number of DMT sorties.
- A **mission element bundling factor** γ, which ranges between 0 and 1, indicating the proportion of the total number of mission element rehearsals that can be *bundled* into mission sorties. For example, if 1000 mission element rehearsals in total are required for combat readiness, and if these can be organized into 800 mission sorties (by fitting many mission elements into a mission sortie), then the bundling factor $\gamma = 0.8$.

Using the above, we can now determine the minimum number of pilots that can be trained for a given training system configuration as follows. Let NTY = TY{$\Sigma_{\kappa=1,45}$AC(k)} denote the actual number of aircraft sorties required in a training system configuration. Similarly, let NTX = TX{ $\Sigma_{\kappa=1,45}$ D_D(k)} denote the actual number of sim sorties required. Therefore, we have:

- MIN_PILOTS_AC = N_AC/{NTY$*\delta_a*\gamma$}
- MIN_PILOTS_D = N_D/{NTX$*\delta_s*\gamma$}

MIN_PILOTS_AC and MIN_PILOTS_D denote the minimum number of pilots that can be trained with the available aircraft and DMT resources respectively, for given levels of the sortie reduction and mission bundling factors. Also, if NTY(γ) < N_AC, then we can simply set $\delta_a = 1$. Similarly, if NTX(γ) < N_D, then we can simply set $\delta_s = 1$. These cases represent situations where the available sorties exceed the training requirements (rare!). We need the reduction factors only when these resources are not adequately available. Further:

- N_AC = {# of aircrafts available}{# of sorties per aircraft per year}
- N_D is computed for 2-ship and 4-ship DMTs as follows:
    - N_2D = 2{# of 2-ship DMTs available}{# of sorties per DMT per year}
    - N_4D = 4{# of 4-ship DMTs available}{# of sorties per DMT per year}

5.2.2.4. Training Level Estimation

This analysis is the converse of the capacity estimation model. In this case, we fix the available resources (such as aircraft and DMT) and estimate the level of practice in the mission elements that can be accomplished if we need to train a give set of pilots in the inexperienced and experienced categories. We summarize this analysis as follows.

- **SME INPUTS**
  - SME assessments IMP, EFAC, INAC, IN2D_AC, IN2D_2D, IN4D_AC, IN4D_4D, EX2D_AC, EX2D_2D, EX4D_AC, EX4D_4D
  - Transfer functions: $\boldsymbol{y}_k = A_k e^{-B_k x_k} + C_k$, where $\boldsymbol{y}_k$ = {# of A/C sorties in mission element k}, $\boldsymbol{x}_k$ ={# of sim sorties in mission element k}
  - Composite Transfer function: TY(P%) = f(TX(P%)), where P% is the step parameter used in the construction of the composite transfer function.
- **USER INPUTS**
  - NPILOTS : Number of pilots to be trained in a year
  - NUM_AC : Number of Aircrafts available
  - AC_SORT: Number of available sorties/aircraft
  - NUM_D : Number of DMTs available
  - D_SORT: Number of available sorties/DMT
  - γ : Mission bundling factor (between 0 and 1), determined from training program
  - n : the number of P% reductions in aircraft time to be employed in selecting an aircraft/sim training system configuration.
- **DECISION PARAMETERS**
  - δ_a and δ_s : These parameters measure the extent to which the required level of training can be fulfilled under the conditions specified in the user specified parametric settings.
  - δ_a = **{NUM_AC}{AC_SORT}/[{NPILOTS}{TY(nP%)}**{$\Sigma_{\kappa=1,45}$ **AC(k)}**{γ}
  - δ_s = **{NUM_D{D_SORT}/[{NPILOTS}{TX(nP%)}**{ $\Sigma_{\kappa=1,45}$ **D_D(k)}**{γ}

In this analysis, the following parametric sensitivity characteristics are important:

1. As n increases (greater use of DMT and less use of aircraft):
   - TY(nP%) decreases (proportion of aircraft sorties)
   - TX(nP%) increases (proportion of sim sorties)
   - Consequently, δ_a increases (the level to which the required aircraft sorties can be fulfilled with the available aircraft sorties)
   - Hence, δ_s decreases (the level to which the required sim sorties can be fulfilled with the available sim sorties)
2. As γ increases (greater bundling leading to fewer sorties to accomplish all training)
   - Both δ_a and δ_s decrease. This indicates that the better we are able to bundle the mission element rehearsals into missions, the greater will be the level to which we can satisfy the sortie requirements for combat readiness, within the available resources.
3. As NUM_AC and/or AC_SORT increase:

- δ_a increases. This is an expected result. BY increasing the resource levels, we can better satisfy the training requirements.

4. As NUM_D and/or D_SORT increase:
   - δ_s increases. This is an expected result. BY increasing the resource levels, we can better satisfy the training requirements. This result is the same as in the case of the aircraft above.

5. As NPILOTS increases:
   - Both δ_a and δ_s decrease. This indicates that greater the training load, the smaller the degrees to which we can train each pilot on both the aircraft and sim, given the resource restrictions.

Finally, if the user wishes to modify the basic SME assessments, he can do so. In this case, the entire analysis, starting from the determination of the transfer functions and the composite function will be repeated with the new set of base data the user might provide. We anticipate this analysis in practice, as there may be users who may not fully agree with the SME assessments. However, we envision only minor changes to the basic SME data from the users.

5.3. Modeling Costs

The costs associated with DMT systems are modeled in terms of recurring and nonrecurring acquisition costs, and direct and indirect operational costs over a time horizon. The nonrecurring acquisition costs include the physical facilities such as training center and database development center. The recurring acquisition costs include simulator hardware and software, IOS, brief/debrief facilities, visual systems, network systems, DMT control station, DMT threat system, DMT data logger system and other software support. These broad components are detailed into specific items in the spreadsheet model. The direct costs include instructional costs and the administration of the training center and database center. The indirect costs pertain to the overheads on training and management. The costs associated with each configuration of DMT systems have been compiled in the spreadsheet. The decision support component of the spreadsheet enables a user to specify the desired levels of practice in aircraft and DMT sorties, training loads, aircraft and DMT resources available and the configuration of the training program in terms of aircraft and DMT sorties to be used. The spreadsheet model estimates the annual prorated costs by taking into account the extensions to aircraft lives due to the reduction in aircraft sorties due to the introduction of DMT in the training environment. The net cash flows over a 15 year period are determined from this analysis, and a net present value of them is calculated at an user specified interest rate. The analyses show a significant positive net present value of the cash flows over this period, which substantially justifies the investments in DMT systems.

5.4. System Architecture

The decision support system has been developed in MS Excel 7.0. The system consists of the following series of linked and integrated worksheets:

- **QUEST1, QUEST2, QUEST3, QUEST4, QUEST5, QUEST6:** These worksheets contain the primary SME data collected from Roadrunner'98, and the data processed from it to generate inputs for the decision support system
- **INPUTS:** This worksheet presents the processed data input from the primary SME data
- **TRANSFERS:** This worksheet calculates the transfer of training functions
- **2SHIP, 4SHIP:** These worksheets perform the parametric sensitivity analyses on the tradeoffs, sortie requirements and composite transfer estimations for training with 2-ship and 4-ship DMT respectively, in conjunction with aircraft.
- **CAPACITY:** This worksheet performs parametric sensitivity analyses on the training capacities of various training system configurations using a *user guided test drive* approach.
- **TR_LEVELS:** This worksheet performs parametric sensitivity analyses on the training levels of various training system configurations for a given set of training loads and training resources available using a *user guided test drive* approach.
- **EFFECTIVENESS:** This worksheet presents the differential transfer effectiveness among aircraft, 2-ship and 4-ship DMT systems using graphical displays.
- **ACQ_COSTS_NR, ACQ_COSTS_RR:** These spreadsheets capture and analyze the nonrecurring and recurring acquisition costs associated with each DMT configuration.
- **DIRECT_COSTS, INDIRECT_COSTS:** These spreadsheets capture and analyze the direct and indirect operational costs associated with each DMT configuration.
- **COST_ANALYSIS:** This worksheet performs the parametric sensitivity analyses on the costs by taking into account all the cost elements and training strategies for each DMT configuration using a *user guided test drive* approach.

The worksheets are sequentially linked. Hence, a user could perform an analysis in a worksheet by changing the parameters in any of the preceding worksheets. The worksheet INPUTS which contains the fundamental processed SME assessments, serves as the root of this sequence.

6. Conclusions

In summary, this research has yielded the following cost - effectiveness assessments leading to a spreadsheet based decision support system for DMT training system evaluation:

- A quantitative comparative training effectiveness analysis of the three training system configurations (No DMT, With 2-Ship DMT, With 4-Ship DMT)

- A quantitative assessment of the effects of training task importance and training effectiveness of the three training systems on the sorties required on these systems for combat readiness under the three configurations.
- A quantitative assessment of the differential training characteristics of the three training systems.
- Estimation of the transfer functions for the mission elements, the composite aircraft-DMT transfer curve, the trainee capacities in each configuration and the costs associated with DMT systems
- Estimation of An overall decision support system for the selection and parametric sensitivity analyses of the various training and cost parameters.

References

1. Braby, R., Charles, J., Sylla, S., Ramesh, R., Willis, R., Hunter, D., "Instructor Considerations in the Design of Optimal Training Devices", U. Army Research Institute Contract No. N61339-86-D-0009-0003, 1988.
2. Derrick, W.L. and Davis, M.S., "Cost-Effectiveness Analysis of the C-130 Aircrew training System", Human Resources Directorate, Aircrew training research Division, Williams Air Force Base, AZ
3. Moor, W.C., "An Empirical Test of a Method for Comparison of Alternative Multiship Aircraft Simulation Systems Utilizing Benefit-Cost Evaluation", AFOSR SREP, 1993.
4. Moor, W.C. and Andrews, D.H., "Benefit-Cost Model for the Evaluation of Simulator Based Multiship Training Alternatives", Armstrong laboratory, Williams AFB, AL-TP-1992-0023, 1992.
5. Mudd, D.H., Pierce, B.J., Greschke, D.J., Kochert, T.J., Surber B. Sell, R. and Gasson, R.D., "Feasibility of CF-18 Multitask trainer", Air Force Research Laboratory, Mesa, AZ, 1997.

RELIABILITY AND VALIDITY TESTING OF THE STUDENT CHARACTERISTICS SCALE

Nancy J. Stone
Associate Professor
Department of Psychology

Creighton University
2500 California Plaza
Omaha, NE 68178

Final Report for:
Summer Faculty Research Program
Air Force Research Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Air Force Research Laboratory

September, 1998

RELIABILITY AND VALIDITY TESTING OF THE
STUDENT CHARACTERISTICS SCALE

Nancy J. Stone
Associate Professor
Department of Psychology
Creighton University

Abstract

Stone (1996) proposed that 16 factors underlie the overlapping research areas of engagement, involvement, and self-regulated learning. Based on this information, she developed and preliminarily tested the Student Characteristics Scale (SCS). Initial results indicated that 13 factors might exist (Stone, 1998). Further data were collected to complete the initial analyses. From this more complete set of data, 12 factors, using 83 of original items, were retained and were similar to the preliminary 13 factors identified. Further analyses using the maximum likelihood estimators indicated that another 23 of the original item should be retained. To resolve issues of an insufficient number of items per factor or low alpha reliabilities, 24 new items were generated. This revised form of the SCS was computerized and distributed to 700 enlisted personnel who had completed technical school training within the past 6 months. This project is still in progress and a draft technical report is in progress. This research effort has led to the recent development of collaborative research projects through the Human-Environment Research Center at the United States Air Force Academy.

RELIABILITY AND VALIDITY TESTING OF THE STUDENT CHARACTERISTICS SCALE

Nancy J. Stone

Engagement (e.g., Finn, Folger, & Cox, 1991; Skinner & Belmont, 1993), involvement (e.g., Reed & Schallert, 1993), and self-regulated learning (e.g., Butler & Winne, 1995; Zimmerman, 1986, 1990) are three areas of research which address the process by which students do or do not acquire knowledge. Although there is some overlap among these three areas of research, there are also some distinctions. After a thorough review of the literatures, Stone (1996) identified the underlying characteristics of these three areas. She also found that these characteristics are related to enhanced learning. Based on the correlates and current measures of engagement, involvement, and self-regulated learning Stone (1996, 1998) proposed, developed, and preliminarily tested the Student Characteristics Scale. This scale is intended to tap the underlying components of student engagement, involvement, and self-regulated. Although preliminary testing indicated that the various factors do exist, further testing is necessary to determine the reliability and validity of this scale.

To understand how the Student Characteristics Scale (SCS) was developed, an overview of the areas of engagement, involvement, and self-regulated learning is presented first. Then, the development of the scale and its preliminary tests are discussed.

Engagement has been defined as an orientation, purpose (Ainley, 1993), and motivation (Lee & Anderson, 1993) to learn. Spanning a continuum from disengaged to engaged, disengaged students tend to be passive in learning, to expend little effort, to give up easily, to be bored, and to be withdrawn from learning (Skinner & Belmont, 1993). Moderate engagement

usually includes a focus on achieving minimal tasks (Ainley, 1993), and being attentive (Finn et al., 1991; Lee & Anderson, 1993), involved in class activities (Lee & Anderson, 1993), and persistent (Finn et al., 1991; Skinner & Belmont, 1993). Fully engaged students tend to focus on achieving a deep understanding of the material (Ainley, 1993), to display initiative (Finn et al., 1991; Skinner & Belmont, 1993), to study/work beyond course requirements (Finn et al., 1991; Lee & Anderson, 1993), to be completely absorbed in one's work whereby time becomes distorted (Goff & Ackerman, 1992), to exhibit intense concentration (Skinner & Belmont, 1993), to display persistence and effort (Finn et al., 1991), and to challenge their abilities (Skinner & Belmont, 1993). Thus, the unique components of engagement were proposed to include intense concentration, intense attention, absorption in a task, and desire to learning thoroughly or beyond the course requirements (Stone, 1996).

Students identified as "involved" have also been described as absorbed in their work, displaying intense concentration, challenging their abilities, and exhibiting positive affect (Reed & Schallert, 1993). Yet, involvement often is associated with job and work involvement (Farrell & Mudrack, 1992). That is, involvement was posited to include one's commitment to learning, perception that learning is important, ownership of learning, and conscientiousness. In a sense, involvement is a work ethic of learning.

Finally, self-regulated learning depicts the learner as autonomous (McCombs & Whisler, 1989), self-regulated or self-monitored (Corno & Mandinach, 1983), resourceful (Zimmerman, 1990), and strategic (McCombs & Whisler, 1989). According to Butler and Winne's (1995) model of self-regulated learning, a self-regulated learner who is given a task will evaluate that task, set goals based on that evaluation, use various strategies to meet those goals, and monitor

activities to assess progress toward those goals. A reinterpretation of the learning process occurs given internal or external feedback about the task, goals, strategies, and outcomes (Butler & Winne, 1995). Therefore, the unique components of self-regulation were proposed to include goal setting or planning, effective strategy use, monitoring of the learning process, and feedback.

Based on Stone's (1996) review, she proposed that 16 factors underlie engagement, involvement, and self-regulated learning. The six proposed engagement factors were interest, attention, absorption, persistence, effort, and desire to learn thoroughly or beyond course requirements. School or learning importance, commitment, conscientiousness, and responsibility of ownership of learning were the four proposed factors of involvement. Finally, the six proposed self-regulated learning factors included goal development, planning, self-concept, strategies, monitoring, and feedback. Based on these hypothesized factors underlying student learning, Stone (1996) created the Student Characteristics Scale adapting items from various scales as well as creating new items.

In a two-part study, Stone (1998) evaluated this proposed scale. Because a total of 134 items resulted from the adaptation of current measurement items and the generation of new items from descriptions and correlations of these constructs, a card sort was implemented before collecting the factor analytic data. The card sorting task revealed card groupings similar to those proposed for the engagement and involvement scales. The card groupings for the proposed self-regulated learning scales were less clear, but some patterns did emerge. Furthermore, 6 redundant items were identified and eliminated. Based on these findings, Stone (1998) developed a questionnaire with the remaining 128 items using a 7-point Likert-type format

(1=never, 7=always). This questionnaire was administered to college students at a medium-sized, private university and a small, private commuter university.

An exploratory factor analysis, using principal components and varimax rotation was used to evaluate the scale. Stone (1998) reported 13 distinguishable factors. These data supported fairly well the factors proposed to underlie engagement, involvement, and self-regulated learning. Additionally, the respective scaled-scores for these factors were correlated with performance measures. All but two of the scaled-scores were significantly correlated with at least one of the performance measures.

Even though Stone's (1998) factor analysis and validity data supported the notion that the factors are distinguishable and that the scaled-scores tap different aspects of academic performance, these data were preliminary. In particular, the alpha reliabilities for several of the scaled scores were below .70. Additionally, some of the scaled-scores included less than 5 items. Thus, further testing of the SCS was conducted before re-testing this scale on a different population. Additional data were collected and analyzed because the initial sample size was smaller than desired. This process is discussed in the Method section.

Method

Participants

United States Air Force enlisted personnel (N=700) who had participated in technical school training within the past 6 months were mailed a computerized form of the SCS. Their training could include any aspect of enlisted military duties from training to be a security police officer to a jet engine mechanic. The personnel surveyed were technicians only, excluding all supervisors.

Materials

To ensure that the initial findings were not affected by a small sample size, additional data from the same population were collected and added to Stone's (1998) data for the exploratory factor analysis. This second analysis included a total of 574 surveys. An exploratory factor analysis, with principal components and varimax rotation was used. These results were similar to the initial findings, but only 12 factors were identified, using 83 items. The 12 factors overlapped fairly well with the proposed factors even though some of the items shifted among the factors. Thus, the factors were named Desire to Learn, Conscientiousness, Resourcefulness, School as Lifestyle, Planning and Monitoring, Responsibility, Self-Concept, Purpose of Education, Initiative, Absorption, Performance Monitoring, and Identification with School. In additional, all but two of the factors correlated with at least one of the performance measures. Only factors III (resourcefulness) and V (planning and monitoring) did not correlate with any of the performance measures.

An additional exploratory factor analysis was conducted because MacCallum (1998) argued that maximum likelihood (ML) as opposed to principal components should be used when determining underlying factors. Although the results were quite similar, the ML method retained additional items. Thus, the corrected version of the SCS included these 83 items identified from the principal components factor analysis and an additional 23 items from the ML analysis. In addition, 24 new items were generated because the alpha reliabilities for the scaled-scores were less than .70 or there were fewer than 5 items loaded on a factor.

A total of 130 items comprised the revised SCS. The scale items were randomized on the survey. The survey was computerized to simplify the data collection process.

Procedure

Computer disks that contained a copy of this survey were distributed to United States Air Force enlisted personnel who recently (within the past 6 months) had completed technical schooling. They were informed that some items refer to education in general, but others refer to specific courses. The Air Force personnel were asked to consider their recent technical school training when items referred to specific aspects of courses or schooling or when the items referred to "my education." Also, they were asked to think of "academic success" as their technical school success and to consider their technical school instructors when items referred to "faculty" or "teachers." The individuals responded to each item using a 7-point scale (1=never, 2= almost never, 3=rarely, 4=sometimes, 5=usually, 6=almost always, 7=always). For clarity, a couple examples were given. Example 1: "My courses are boring." If you disagree with this statement and think that your courses are NEVER boring, you would respond with a 1=never. Example 2: "It is more important that I study rather than go out with friends." If you think that USUALLY it is more important to study rather than go out with friends, you would respond with a 5=usually. The participants were informed that there were no correct answers and they were thanked for participating.

Results

As indicated earlier, this project is still in progress, therefore there are no data to report. Nevertheless, the methods of analysis that will be used are discussed.

A confirmatory factor analysis will be used to test the proposed model. After the model is tested, scaled-scores for each factor will be calculated and compared to various outcome measures. In particular, the scaled-scores will be compared with the respondents' educational level; technical school scores; basic military training scores; mechanical, administrative, general, and electrical scores derived from the Armed Services Vocational Aptitude Battery (ASVAB); and the Armed Forces Qualification Test (AFQT) score, also a composite of various ASVAB scores. These measures will be used to validate the scales of the SCS.

Discussion

In summary, the SCS was revised from previous data. This included the generation of new items to add to the items retained from the exploratory factor analyses. Because the data are still being collected, no conclusive results may be discussed. Once all data have been collected, the data will be reported in final form. In particular, a draft technical report is in progress.

In addition to refining the SCS this summer, the summer project has also led to the development of collaborative projects with Col. David Porter in the Department of Behavioral Sciences and Leadership at the United State Air Force Academy (USAFA). This collaboration evolved with the assistance of Dr. James Miller in the Human-Environmental Research Center (HERC) at the USAFA. Furthermore, collaborative projects have been discussed with Dr. Kieth Carlson and Dr. Steve Jones, both members of the Department of Behavioral Sciences and Leadership at the USAFA.

References

Ainley, M. D. (1993). Styles of engagement with learning: Multidimensional assessment of their relationship with strategy use and school achievement. Journal of Educational Psychology, 85, 395-405.

Butler, D. L. & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. Review of Educational Research, 65, 245-281.

Corno, L. & Mandinach, E. B. (1983). The role of cognitive engagement in classroom learning and motivation. Educational Psychologist, 18, 88-108.

Farrell, G. M. & Mudrack, P. E. (1992). Academic involvement and the nontraditional student. Psychological Reports, 71, 707-713.

Finn, J. D., Folger, J., & Cox, D. (1991). Measuring participation among elementary grade students. Educational and Psychological Measurement, 51, 393-402.

Goff, M. & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of typical intellectual engagement. Journal of Educational Psychology, 84, 537-552.

Lee, O. & Anderson, C. W. (1993). Task engagement and conceptual change in middle school science classrooms. American Educational Research Journal, 30, 585-610.

MacCallum, R. (1998). Comentary on quantitative methods in I-O research. The Industrial-Organizational Psychologist, 35(4), 18-30.

McCombs, B. L. & Whisler, J. S. (1989). The role of affective variables in autonomous learning. Educational Psychologist, 24, 277-306.

Reed, J. H. & Schallert, D. L. (1993). The nature of involvement in academic discourse tasks. Journal of Educational Psychology, 85, 253-266.

Skinner, E. A. & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. Journal of Educational Psychology, 85, 571-581.

Stone, N. J. (1996). Engagement, involvement, and self-regulated learning: Construct and measurement development to assess achievement and calibration. Unpublished manuscript, Creighton University.

Stone, N. J. (1998). Evaluation of a scale designed to measure the underlying constructs of engagement, involvement, and self-regulated learning. Unpublished manuscript, Creighton University.

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? Contemporary Educational Psychology, 11, 307-313.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. Educational psychology, 25, 3-17.

PRELIMINARY DECISION ANALYSIS OF THE DATA EXPLOITATION, MISSION PLANNING, AND COMMUNICATION (DEMPC) SYSTEM OF THE PREDATOR UNMANNED AERIAL VEHICLE (UAV)

Robin D. Thomas
Assistant Professor
Department of Psychology

Miami University
Oxford, Ohio

Final Report for:
Summer Faculty Research Program
The Air Force Research Laboratory
Brooks AFB, Texas

Sponsored by
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

The Air Force Research Laboratory

September 1998

PRELIMINARY DECISION ANALYSIS OF THE DATA EXPLOITATION, MISSION PLANNING, AND COMMUNICATION (DEMPC) SYSTEM OF THE PREDATOR UNMANNED AERIAL VEHICLE (UAV)

Robin D. Thomas
Assistant Professor
Department of Psychology
Miami University
Oxford, Ohio

Abstract

A decision analysis of the DEMPC mission planning and replanning functions was considered. The DEMPC is a component of one of the three subsystems in a Predator UAV Ground Control Station. The job of this operator is to determine the best path through a region of territory that the Predator UAV will navigate in order to image (i.e, take pictures or radar images of) various locations, or targets, to satisfy reconnaissance objectives. Factors such as relative location, wind direction and speed, cloud cover, forbidden zones, type of target, etc., all influence the decision strategy likely to be employed by this operator. The present paper explores the possible mathematical structure of an optimal strategy to perform this task and discusses the applicability of such analyses to actual human performance. A preliminary empirical investigation is outlined which is designed to reveal decision strategies and sensitivities to the various factors exhibited by individuals faced with mission planning. The long term goal is to develop a cognitive performance model of the DEMPC task that can be integrated into a Predator - like system serving as an intelligent agent.

PRELIMINARY DECISION ANALYSIS OF THE DATA EXPLOITATION, MISSION PLANNING, AND COMMUNICATION (DEMPC) SYSTEM OF THE PREDATOR UNMANNED AERIAL VEHICLE (UAV)

Robin D. Thomas

Unmanned Aerial Vehicles (UAVs) have become an increasingly important tool in the concept of a modern armed forces. They are particularly well suited for safely supporting military forces by monitoring civilian activities, troop locations, artillery positions, garrison activities, and compliance with agreements. The signing of the Dayton Peace Accord ending the violence in Bosnia is due in part to the successful deployment of the Air Force UAV, the Predator system, in both surveillance and targeting support of operations Provide Promise and Deliberate Force (Israel, 1997). In furthering the development of the Predator UAV system, research into the human factors and cognitive performance requirements has begun.

The Predator is a medium altitude endurance UAV that is currently designed to generate live imagery and imagery-derived intelligence for use in a variety of military tasks. This goal is accomplished by a system that is organized into several components collectively referred to as the Ground Control Station (GCS) of which the most important elements are the Air Vehicle Operator (AVO) who pilots the aircraft remotely, the Payload Operator (PLO) who operates the onboard daylight camera or infrared sensors also remotely, and the Data Exploitation, Mission Planning, and Communication (DEMPC) operator, who plans the mission beforehand and replans it during flight as new targets (i.e., locations to be imaged) are assigned or weather conditions change. In effect, the DEMPC is the 'local command and control center' of a Predator mission, though significant input is given from a variety of other upper echelons (Hall &

Gugerty, 1997). This paper describes preliminary investigations into the cognitive processes underlying the planning and replanning dynamics of the DEMPC. The long term goal is to develop a cognitive performance model of the DEMPC task that can be integrated into a Predator - like system serving as an intelligent agent.

The DEMPC mission planning task

The mission planning job of the DEMPC can be divided into a two parts: preflight and during flight. Before a typical Predator mission begins, upper echelons assign a list of targets in the form of latitude/longitude and often altitude coordinates together with the reconnaissance objectives for each. These targets lie in predefined areas within which the UAV must remain during flight referred to as Restricted Operating Zones (ROZ), which in Bosnia were approximately 30 miles by 30 miles. The UAV is also constrained to visit the targets to be imaged during an allotted time period.[1] Deviations from these constraints require clearance from various authorities. Targets differ in terms of priority, that is, their need to be imaged ranging from low (nice if one can get to it safely) to very high (a must have). Examples of types of targets often imaged include hangars, aircraft on an airfield, or intersections of streets. Images actually taken need to be of high enough quality to answer questions about the target such as what is the number of planes in the hanger at a particular time or how many troops are gathering at a given intersection. Prior to the execution of a mission, the DEMPC, using a computer workstation[2], will plot a path through the ROZ box that allows the targets to be imaged given the time allotted and fuel requirements of the craft. Additional information that is available to the DEMPC for planning purposes are the required altitudes for imaging and the prevailing weather (especially wind). Figure 1 shows a diagram, referred to as a cartoon, containing ROZ boxes and

other information that the DEMPC uses to aid the planning of the mission.

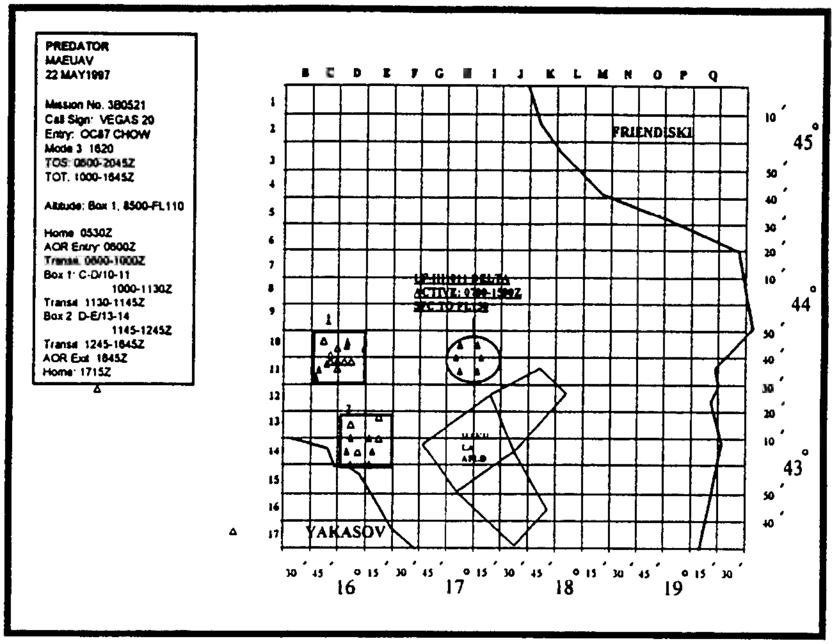


Figure 1. A cartoon depicting several restricted operating zones (ROZ) in which clusters of targets are located. Information about the locations appear on the left. From Hall & Gugerty

The DEMPC's preflight mission plan then is to find an efficient path through the ROZ box(es) that allows the reconnaissance objectives for multiple targets to be satisfied given the constraints of time, known weather conditions, and the aircraft performance. During flight, very often new targets are added, call ad hocs, and are almost always of the highest priority. In addition, as typical missions can take several hours, weather conditions change at the sites. For example, wind direction and speed throughout the ROZ box can be different than forecast which would affect time and fuel consumptions. Most importantly, cloud cover at the site may occlude targets and prevent the use of the daylight and infrared cameras. There is synthetic aperture radar (SAR) capabilities that allow radar imaging of certain types of targets which are cloud occluded

but is not good for imaging areas within cities or are otherwise in densely populated locations. The challenge for the DEMPC during flight is to balance these changing conditions when replanning flight paths. All of the constraints which existed preflight are still in effect of course during flight.

Concepts of Optimal Mission Planning

The challenge for modeling the planning and dynamic replanning task is to understand exactly what the various dimensions along which the DEMPC represents the problem and to identify how those are combined to form an overall concept of efficiency that generates a route. The response to be generated is the path, both pre- and during flight, through the ROZ box. Formally, this is an optimization problem. The preflight planning is static while the inflight is a dynamic problem.

In the steps of building an intelligent agent to accomplish this task, one of two approaches could be adopted. First, the cognitive processes employed by experience DEMPCs would serve as the model. This is the expert systems approach. Alternatively, if the problem could be objectively stated, that is, the concept of efficiency be rigorously defined in terms of all of the various costs and benefits, a classical optimization algorithm may be found which simply maximizes the efficiency by choice of a path through N points. This latter, engineering solution while guaranteeing an objectively best choice, may be very difficult in practice as the computational difficulties for this sort of problem are immense. For example, if the goal were to find the shortest route among N targets (ignoring differing priorities, aircraft performance issues, weather, etc.) then any reasonable algorithm proposed for this classic 'traveling salesman' optimization problem would suffice[3]. It is known though, that no one has discovered a solution

that does not show an exponential growth of computer run time with increasing the number of points. Consequently, if employed in real situations, the computational time required by the algorithm may diminish its usefulness. Another consideration is that true optimality may not be required as long as the result is good enough. In that case, heuristic searches through possible paths may be sufficient. Nonetheless, some insights into the problem facing a DEMPC can be achieved by examining the mathematical structure of the task prior to proposing solutions that are actually used. The task for the DEMPC is also complicated by the issues of differing target priorities and the constraints introduced by weather and the ROZ borders becoming something of a weighted traveling salesman problem.

Optimization problems can be broken down into those that have one or more than one criteria. A one-criterion problem maps all of the external and internal variables into a one-valued function that is to be maximized or minimized subject to some constraints. The traveling salesman problem is a one-criterion problem with distance as the criterion to be minimized as a function of routes subject to the constraint of visiting each city only once. If other attributes are included in the problem such as wind direction and speed then other criteria besides distance may become important such as fuel used. The problem could then be stated as a two-criterion problem: simultaneously minimize distance and minimize fuel as a function of route (and wind) subject to the constraint of visiting each city exactly once. Alternatively, one could construct a one-criterion object, called *efficiency* that is a function of both distance and fuel and maximize this new function. This function may be additive or multiplicative or something else depending on the goals of the decision maker and the interdependence of the attributes of the decision (Clemen, 1996). When an individual DEMPC is faced with planning routes, it would be of

interest to reveal what type of optimization function describes his or her performance.

An algorithm that may be used to determine the objectively best path can be summarized by the following steps:

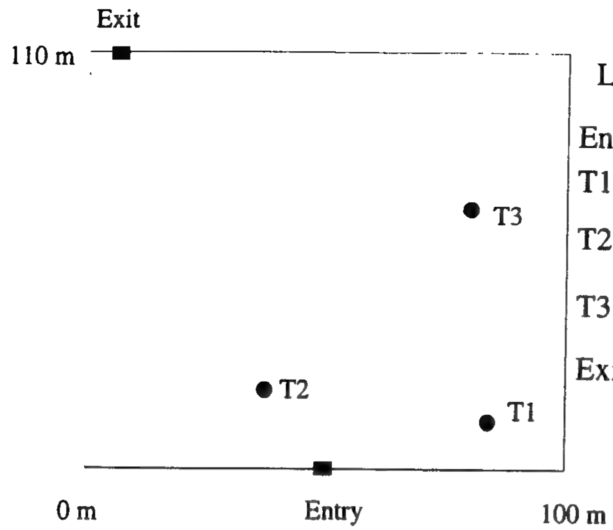
1. There are n! paths through all n points in the ROZ
2. For each path compute:
   a. Total fuel consumed, FC
   b. Total time consumed, TC
   c. Expected Priority, EP = $\sum_{i=1}^{n!} (\textit{Imaging Probability})_i \, x \, \textit{Priority}_i$
3. Reject all paths whose FC > maximum fuel available or TC > maximum time allowed in the ROZ
4. Optimize:
   i. If one path remains after Step 3, choose it as mission plan.
   ii. If no paths remain after Step 3, list all possible paths through n - 1 targets and return to Step 1.
   iii. If more than one path remains after Step 3, select "best" one by the desired optimization strategy

For Step 4iii, as mentioned earlier several possible formulations are admissible. The choice depends on the propensities of the decision maker and special features of the situation. One could use an additive combination rule such as

$$E = \alpha EP - \beta TC - \gamma FC \qquad (1)$$

where α, β , and γ (all > 0) are parameters reflecting the relative importance of each attribute of the decision problem. For a given set of parameters, one would find the path such that E is maximum. In the cases in which two of the parameters are zero, then the problem reduces down to a one-criterion optimization of either choosing the path, ρ_i, such that TC is minimum; or the path, ρ_j, such that FC is minimum; or the path, ρ_k, such that EP is maximum. If all n targets can be visited in the time allotted the EP is constant across paths unless imaging probability changes as a function of time. One example in which imaging probability could change as a function of time is if the UAV is increasingly exposed to engine failures or enemy threats as time progresses. Finally, the addition of elements of a production system may be necessary to capture the fact that certain paths lead the plane into forbidden zones or to altitudes that are impossible to achieve given the size of the engine. Production rules can be incorporated as checks performed on paths prior to the computation of an objective function value. Paths that fail such rules can, a priori, be rejected.

To illustrate this optimization algorithm, consider the following, highly simplified, situation. Figure 2 shows a cartoon describing three target locations in a ROZ with the entry and exit locations along with additional information regarding wind direction and speed, target priorities, fuel flow, and image likelihood for each target given cloud cover over the ROZ.



Locations		Priority	Image Prob.	Wind
Entry (50,0)				330 dg/ 30m/hr
T1	(90,10)	1	.8	
T2	(40,30)	2	.3	
T3	(5,110)	3	.5	
Exit (5, 110)				

Fuel flow rate: 18 lbs/hour

Figure 2. A possible ROZ box for a simple Predator mission illustrating an optimization algorithm for mission planning.

For each of the six (3!) possible paths through the three targets, time required, fuel consumed, and distance traveled are computed and listed in Table 1. Notice that in terms of distance alone, path ρ_3 is best. But when wind is taken into account, path ρ_2 is best for time and fuel. Time and fuel may also be decoupled, though they will always remain interdependent, if altitude climbs are required for some legs and not others. This is because fuel consumption is punished considerably while travel time is almost unaffected.[4] In this case, the most efficient path would be determined by the relative weighting of time versus fuel (e.g., if β were much greater than γ in Eq. 1). Without altitude climbs, fuel is a linear function of time so that these dimensions are actually only one.

Table 1. Possible Paths through the ROZ with Time, Fuel, and Distance consumed for each.

Path	Time (hours)	Fuel (lbs)	Distance (miles)
ρ_1 T1,T2,T3	6.38	114.82	248.58
ρ_2 T1,T3,T2	5.89	106.07	235.20
ρ_3 T2,T1,T3	6.06	109.11	232.41
ρ_4 T2,T3,T1	6.72	120.93	269.47
ρ_5 T3,T1,T2	6.47	116.38	268.80
ρ_6 T3,T2,T1	7.56	136.95	312.72

Suppose that the UAV only had 5.5 hours available for its mission. No path can be completed in less than 5.5 hours so a target will have to be dropped from the list. At this point, image probability and target priority emerge as important determinants of planning. Because the third target to be visited is determined once the order of the first two are selected, there are again six possible two-target paths through the ROZ - the same as before. Because FC = TC x fuel flow, we can replace Eq. 1 with

$$E = \alpha EP - \nu TC \qquad (2)$$

Table 2 gives new distance, time, and fuel statistics for the two-target paths as well as the now relevant expected priority. In this situation, paths r_3 and r_4 yield the highest expected priority whereas path r_1 is best in terms of time and fuel used and r_5 appears to be a good choice as well.

Table 2. Expected priority, time, fuel, and distance for the possible two-target paths.

Path	Exp. Priority	Time (hours)	Fuel (lbs)	Distance (miles)
r_1 T1, T2	1.4	4.95	89.10	182.40
r_2 T2, T1	1.4	5.77	103.97	216.72
r_3 T1, T3	2.3	5.06	91.16	188.17
r_4 T3, T1	2.3	6.29	113.29	258.86
r_5 T2, T3	2.1	4.99	89.80	185.12
r_6 T3, T2	2.1	5.39	97.07	214.90

Depending on the values of α and ν, different paths could emerge as best. In Table 3 are values of efficiency from Eq. 2 for different values of α relative to ν illustrating the effect of the criterion weighting. The best path would be the one with the highest efficiency.

Table 3. Efficiency of three candidate two-target paths for different α with $\nu = 1$.

Path	$\alpha = \nu$	$\alpha = 5\nu$	$\alpha = 1/5\ \nu$	$\alpha = 1/100\ \nu$
r_1	-3.55	2.05	-4.67	-4.94
r_3	-2.76	6.44	-4.60	-5.04
r_5	-2.88	5.51	-4.57	-4.97
Best	r_3	r_3	r_5	r_1

In the dynamic replanning situation, events occur that require a new path to be selected. For example, command and control may request that the UAV visit an additional set of targets referred to as ad hocs. These targets are often of very high priority (Hall & Gugerty, 1997). Online optimal responding is possible if one reapplies the above analyses at different points that require replanning in the execution of the mission. The overall path that is actually traversed may not be the globally best, a posteriori, but will be locally best at each stage in time.

DEMPC as a Decision Maker

The above discussion concerns the ideal decision maker or a computer algorithm that could be implemented to serve as an intelligent agent performing the task. This algorithm would be implementable only if the correct objective function is known (e.g., additive versus multiplicative) along with the various weighting parameters (e.g., α, β, and γ in Eq. 1). In modeling actual human performance, one class of decision models, based on utility theory,

argues that decisions made by individuals are accomplished largely in the same manner as an optimal responder but that the criteria (FC, TC, and priority) are mapped into subjective "utilities" prior to the application of the combination rule of the objective function. These utilities reflect the subjective value or desirability attached to the attribute. In addition, any event probabilities, such as imageability, are also subjectively determined. That is, the decision maker has a desirability or utility attached to fuel consumption, time consumption, and priority which are all unobservable and may need to estimate important probabilities from experience as well before he or she can determine a best path. Individuals will differ in their assignment of utility and probability to the various constructs but once assigned their behavior is consistent with the maximizing of expected utility according to this approach. Though the quantities in such computation are unobservable, the structure of the decision algorithm will imply systematic relations hold across the decision maker's choices. For example, if path A is preferred to path B, and path B is preferred to path C, then path A should be preferred to path C. Experiments are then constructed which test for these invariances to assess the performance of the model. It is also possible to formulate exact functions for Eq. 1 and estimate the parameters, (i.e., fit a model) from the choice probabilities for various paths and examine how well the model agrees with the data.

It is often observed that experts performing a task require vastly smaller amounts of computational time and memory storage than equivalent computer algorithms based on traditional utility theory constructs (witness the classic challenges between IBM's Big Blue and a human chess master) yet somehow they perform at or nearly at optimal levels. The research into problem solving strategies and heuristics has attempted to reveal the nature of expert cognition.

This research suggests that experts differ from novices not in the decision process they employ but rather in their representation of the elements of the problem (Holyoak, 1995). The bottom line is that experts categorize the situation more accurately than nonexperts. Once the situation is categorized, the response chosen is automatic. Realistic or naturalistic approaches to decision making, incorporate these ideas in representation based models of choice (Zsambok & Klein, 1997). The burden of computation, it is argued, is not in the formulation and computation of the objective function, rather, it is in the perception and classification of the available data into an appropriate situation. Again, once the situation is recognized, the action to be taken is deterministically specified. This approach has been successfully applied to the understanding of command and control decisions in typical military situations. While clearly promising, it is not readily apparent how to apply this approach to the continuous response situation demanded of a DEMPC but there may be possibilities if further study is undertaken.

Preliminary Empirical Investigation of Factors Affecting Mission Planning and Replanning

As a beginning evaluation of the applicability of decision theory models to the task of the DEMPC it is desirable to augment the cognitive task analyses which were based on interviewing actual DEMPCs (Hall & Gugerty, 1997) with a body of actual decision performance data collected on a DEMPC-like task. With that in mind, we are planning an experiment using a simulated DEMPC station in which various mission scenarios will be given to a participant for consideration.[5] This individual will be asked to preplan a route through a ROZ in order to collect images of a given list of targets and then, as the mission unfolds, replan if any new ad hoc targets appear or when changes in weather or image probabilities occur (e.g., cloud cover is discovered, etc.). They might also have to replan if it is discovered that the path they had chosen leads them

to run out of time or fuel. The scenario faced by the participants will be more complex than the simplified example outlined above in several respects. First, the number of targets to be imaged will be more numerous and they will be of differing types (hangars, tanks, people, etc.). The type of target will influence the altitude requirements (e.g., the UAV will need to be relatively low to image the inside of a hangar to determine if it contains a plane), the kind of payload camera that can be used (e.g., daylight optical versus radar), and how difficult it will be to get a clear view. This latter characteristic will impact the time and fuel consumed during the UAV's loiter around the target. Another difference between the simplified situation and a more realistic one is that the DEMPC actually will need to plan a series of points referred to as collection points at which the onboard cameras are able to image various targets. All that is necessary for a target to be imageable (provided no cloud cover exits or other local characteristics obscure the target, etc.) is that the UAV be within a certain radius, usually two miles, of the target. A consequence of this is that in some cases, it is possible to collect more than one target at a given collection point. Many of these factors will be incorporated in the computer simulation of this task. The goal of the study is to systematically sweep out combinations of target types, altitude requirements, relative locations, priorities, weather, ad hoc probabilities, time and/or fuel constraints, etc. so that tradeoffs among them have to be made by the decision maker. His or her responses to a sequence of these events will help to reveal the relative weightings of different factors and possibly how those factors are combined in producing a route. The goal for the participant will be to image as many targets as possible given the time and fuel available. Initially, a scoring algorithm which takes into account priority of the targets successfully imaged, fuel and time consumed, and any incursions into forbidden zones will be implemented to motivate the

participant in the task. It should be noted, however, the scoring algorithm is itself an objective function which the participant is asked to maximize. Unfortunately, if one is interested in revealing the nature of someone's objective function (together with the relative weightings, etc.), the use of a scoring system will eliminate this capability. The data will reveal how optimal, or suboptimal, given the scoring algorithm, the individual is and perhaps in what ways they deviate from optimality. When real DEMPCs can be used, the scoring algorithm should be dropped and only feedback about the number of targets imaged and fuel/time statistics will be provided.

Method

Participants. Student pilots from college in San Antonio, TX will provide the initial data with actual DEMPCs from the U.S. Air Force serving in a follow up study.

Procedure. A mission planning and preplanning task are simulated on a PC that will both run the experiment and record the participants' responses. The interface with which the participant will interact consists of three different screens each of which can be visited from any other screen. Figure 4 displays two of the three screens, the mission planner and mission monitor. The target information screen simply lists the target locations, priorities, and type of target and whether one can use the synthetic aperture radar to image it, and the time restrictions for being in the ROZ box. The mission planner screen allows the individual to plot out a series of collection points, using the mouse, that are used by the simulation to image, if possible, the targets in the ROZ. For each collection point plotted, the individual will also input the altitude at which he or she wants the UAV to be. Like an actual mission planner in the DEMPC workstation, this screen also provides projections for time and fuel consumption given the collection point location (with altitudes) and wind conditions. If, during a mission run, the remaining mission needs to be

replanned, it is done on this same screen so the fuel used/remaining and time statistics are displayed. The mission monitor screen displays the results of the execution of the mission plan: time and fuel used and whether the targets within range of the collection points have been successfully imaged. If a failure occurs, the participant is prompted to either continue to loiter in the vicinity of the target or abandon it and continue with the rest of the mission plan. At any point in time, the individual is permitted to replan the mission by returning to the mission planner screen.

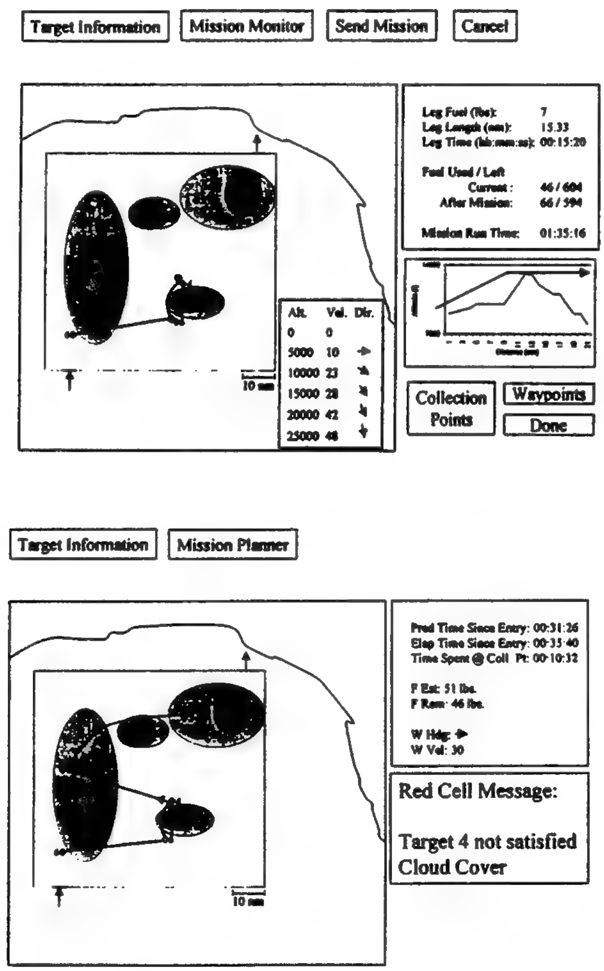


Figure 4. (Top) A mission planner screen display used in the computer simulated DEMPC task. (Bottom) The mission monitor screen displaying mission outcomes and statistics.

Design. The trial scenarios are constructed from a sampling of several factors thought to be important in the planning process. These are categorized as quadrant location in the ROZ, density of targets in the quadrant (1 or 4), en-route resources consumed by the UAV within a quadrant (high or low implemented by manipulating wind, geographic features, locations of forbidden zones, etc.), loitering (at target) resources consumed by UAV (cloud cover, natural or man-made obstructions, special imaging requirements of target), and target priority (high or low). Trial sequences are constructed by randomly sampling from the orthogonal combinations of these characteristics across each quadrant and/or target. The types of targets are weapons piles, gatherings of troops, missile launching sites, vehicles, tanks, and planes on runways or in hangars.

Other Issues: Individual differences in risk perception and acceptance

It is known that individuals differ in their propensity to accept risky choices (Hurwitz, 1998) with some individuals averse to risk and others seemingly seeking risk. They also differ in their perception of actual risk in the decision problem. Hurwitz (1998) has devised assessment techniques that can classify individuals along risk acceptance and risk perceptiveness continua. It is of interest to ascertain whether these individual differences in risk behavior appear in the DEMPC mission planning choices and if they have any impact on actual mission success. The identification of differences in risk attitudes and their consequences would be important for personnel selection and/or training of UAV DEMPC operators.

References

Clemen, R. T. (1996). *Making hard decisions.* Boston: PWS-Kent.

Hall, E. & Gugerty, L (1997). Predator operations cognitive task analysis results. Briefing given to Armstrong Laboratory, Aircrew Training Division (AL/HRA), Mesa, AZ.

Holyoak, K. J. (1995). Problem solving. In E. E. Smith and D. N. Osherson (Eds.), *Thinking: An invitation to cognitive science* (Vol. 3). Cambridge, MA: MIT Press.

Hurwitz, J. (1998). Modeling time-pressured risky decision-making. Proceedings of the Association for Information Systems.

Israel, K. R. (1997). Modeling and simulation employed in the predator unmanned aerial vehicle program. Report to the Defense Airborne Reconnaissance Office.

Zsambock, C. E., & Klein, G. (1997). *Naturalistic decision making.* Hillsdale, NJ: Erlbaum.

Footnotes

1. Nearly all of what is known about Predator performance was obtained during its deployment in Bosnia. Future missions may deviate in detail from the 'typical' one described but the overall task decomposition is likely to be essentially unchanged.

2. The computer station is called the Data Exploitation, Mission Planning, and Communication System (DEMPC). In common usage, however, the person operating the station has inherited the name as well. The actual workstation has a great many hardware and software details of which only the essential elements are featured in the present analysis.

3. This problem involves taking locations of a number of cities and attempting to determine the shortest route that passes through each of them only once. No method has been proposed other than brute-force search through all possible routes selecting the shortest one.

4. This assumes that velocity along the flight path is constant which, in the face of gravity, would necessitate an increase in thrust consuming more fuel. A typical angle of climb is less than 5 degrees which would only affect the horizontal velocity by less than one-two percent, and hence the time of flight is relatively unchanged.

5. This work is being done in collaboration with Dr. Joshua Hurwitz of the AFRL, Brooks, AFB, Texas.

THE EFFECT OF REPEATED MEASUREMENTS ON THE VARIANCE OF THE ESTIMATE OF THE HALF LIFE OF DIOXIN IN THE AIR FORCE HEALTH STUDY

Ram C. Tripathi
Professor
Division of Mathematics and Statistics

University of Texas at San Antonio
6900 N. Loop 1604 West
San Antonio, TX 78249

Final Report for:
Summer Faculty Research Program
Air Force Research Laboratory
Brooks Air Force Base

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Air Force Research Laboratory (AFRL/HEDB)
Brooks Air Force Base, Texas

September 1998

THE EFFECT OF REPEATED MEASUREMENTS ON THE VARIANCE OF THE ESTIMATE OF THE HALF LIFE OF DIOXIN IN THE AIR FORCE HEALTH STUDY

Ram C. Tripathi
Professor
Division of Mathematics and Statistics
University of Texas at San Antonio

Abstract

The half-life of dioxin in the veterans of Operation Ranch Hand depends on its decay rate, denoted by λ. Estimates of the decay rate have been obtained by postulating a log-linear model relating logarithm of dioxin to time since exposure and other covariates of interest and estimating the coefficients by the method of least-squares. So far, four measurements per subject on dioxin, time since exposure, and other covariates are available, with future measurements planned. The postulated model depends on the law of first order kinetics and holds only when the dioxin measurements for each subject are beyond a threshold (taken here as 10 parts per trillion (ppt)). Due to this selection criterion, the least-squares estimate of λ is biased. It can be made unbiased by further selecting the subjects whose dioxin measurements lie above a tilted line. This results in a loss of subjects. As more measurements on each subject become available, the variance of the estimate should decrease, however, that need not be the case. As one obtains more measurements on each subject, those with dioxin measurements lower than the threshold of 10 ppt are excluded from the study. Thus the gain realized by additional measurements may be offset by the loss of subjects. In this report variance of the corrected unbiased estimate of the decay rate is compared for various values of k, the number of repeated measurements. This study will be helpful in policy decisions regarding continuation of the dioxin half-life study.

THE EFFECT OF REPEATED MEASUREMENTS ON THE VARIANCE OF THE ESTIMATE OF THE HALF-LIFE OF DIOXIN IN THE AIR FORCE HEALTH STUDY

Ram C. Tripathi

Introduction

The half-life of dioxin in veterans of Operation Ranch Hand depends on its decay rate, denoted by λ. Estimates of the decay rate have been obtained by postulating a log-linear model relating the logarithm of dioxin to time since exposure and other covariates of interest and estimating the coefficients using the method of least-squares. So far, four measurements per subject on dioxin, time since exposure, and other covariates are available, with future measurements planned. The postulated linear model depends on the law of first order kinetics and holds only when the dioxin measurements on every subject are beyond a certain threshold (taken here as 10 parts per trillion (ppt)). Due to this selection criterion, the least-squares estimate of λ is biased. It can be made unbiased by further selecting the subjects whose dioxin measurements lie above a tilted line (Michalek, et. al. (1998)). This results in a loss of subjects. As one collects more measurements on each subject, the variance of the estimates of λ should decrease, however, this need not be the case. As one obtains more measurements on each subject, the future dioxin measurements may be lower than the threshold of 10 ppt due to continuous decay. This results in a further loss of subjects included in the study. Thus, the gain realized by additional measurements may be offset by the loss of subjects with dioxin below the threshold. Hence, it is important to study how variance of the estimate of the decay rate is affected by additional measurements.

In this report, we obtain closed form expressions for the least-squares estimate of the decay rate for $k = 3, 4, 5$. The expressions for the variance of this estimate, corrected for selection bias, are also derived. With parameter estimates obtained from the Ranch Hand data for $k = 3$ measurements, the variance of the resulting

estimate of the decay rate is computed when the log(dioxin) measurements for each subject have a truncated k-variate normal distribution. This requires the evaluation of truncated multivariate normal integrals. FORTRAN programs using multivariate quadrature subroutines from the NAG and IMSL libraries have been written for computing the variances for $k = 3, 4, 5$. The results are compared.

Model Development and Estimation

Assume that k observations are taken for each of n subjects. These subjects were exposed to a contaminant that produced an elevation in the body burden greater than a background level. Let C_0 denote the initial (unknown) concentration and C_t the concentration t years after exposure. Then a first-order kinetic model

$$C_t = C_0 e^{-\lambda t} \tag{1}$$

holds in the subjects with body burden above a threshold c, where λ denotes the unknown decay rate. Based on equation (1), the population half-life of the contaminant is given by $t_{\frac{1}{2}} = \frac{log(2)}{\lambda}$. By taking the natural logarithm of equation (1), we obtain

$$log(C_t) = log(C_0) - \lambda t. \tag{2}$$

Equation (2) can be regarded as a motivation for a linear regression model with repeated measurements incorporating subject effects. Let y_{ij} denote the log(dioxin) of the j^{th} measurement on the i^{th} subject taken at time t_{ij}. Then, the linear model, accounting for subject effects, is given by

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \tau_i + \varepsilon_{ij}, \tag{3}$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, k$, where $\sum_{i=1}^{n} \tau_i = 0$. Here $\beta_1 = -\lambda$, and ε_{ij} denotes normal error with 0 mean.

The inclusion of subjects with $y_{ij} > log(10)$ in the study causes left truncation, and the least-squares estimates of the parameters are biased. In order to correct the estimate of β_1 for the bias and in order to obtain the variance of the resulting

estimate, it is necessary to obtain its explicit form. In what follows, we give the expression for the least-squares estimate of β_1 and suggest how to make it unbiased by appropriately choosing the subjects under study (Michalek, et. al. (1998)). For these developments, we need the following notations.

Notations: For $i = 1, 2, \ldots, n$

•$\boldsymbol{y}_i$ denotes $k-$dimensional vector of observations for subject i, $\boldsymbol{y}_i^t = (y_{i1}, y_{i2}, \ldots, y_{ik})$,

•$\boldsymbol{t}_i$ denotes $k-$dimensional vector of times for subject i, $\boldsymbol{t}_i^t = (t_{i1}, t_{i2}, \ldots, t_{ik})$,

•Φ denotes the covariance matrix of $\boldsymbol{y}_i$,

•$\boldsymbol{\alpha} = \mathbf{1}^t\Phi^{-1}\mathbf{1}$, where $\mathbf{1}$ denotes the $k-$dimensional column vector with all elements equal to 1,

•$t_i^* = \mathbf{1}^t\Phi^{-1}\boldsymbol{t}_i$,

•$\bar{\boldsymbol{t}} = n^{-1}\sum_{i=1}^n \boldsymbol{t}_i$,

•$\bar{t}^* = n^{-1}\sum_{i=1}^n t_i^*$,

•$S^2 = \frac{1}{n}\sum_{i=1}^n(\boldsymbol{t}_i - \bar{\boldsymbol{t}})\Phi^{-1}(\boldsymbol{t}_i - \bar{\boldsymbol{t}})$,

•$CM = \boldsymbol{t}^t\Phi^{-1}\bar{\boldsymbol{t}}$.

Then, the least-squares estimate of β_1 is given by (see Tripathi(1998))

$$\hat{\beta}_1 = \frac{n}{[D - n^2 s^2]}\left[\sum_{i=1}^n \boldsymbol{a}_i^t \boldsymbol{y}_i\right],$$

where $D = n^2[\alpha^2(S^2 + CM) - \bar{t}^2]$, $s^2 = \frac{1}{n}\sum_{i=1}^n(t_i^* - \bar{t}^*)^2$ and

$$\boldsymbol{a}_i^t = [\boldsymbol{\alpha}^t - \mathbf{1}^t\boldsymbol{\Phi}^{-1}\mathbf{1}^t]\boldsymbol{\Phi}^{-1} = [a_i[1], a_i[2], \ldots, a_i[k]],$$

with $\sum_{j=1}^k a_i[j] = 0$ or $\sum_{j=1}^{k-1} a_i[j] = -a_i[k]$. Hence,

$$\hat{\beta}_1 = \frac{n}{[D - n^2 s^2]}\left[\sum_{i=1}^n\sum_{j=1}^n (a_i[j] - a_i[k])(y_{ij} - y_{ik})\right].$$

Since the $y'_{ij}s$ appear in $\hat{\beta}_1$ only as differences, it will be shown below that $\hat{\beta}_1$ can be made unbiased by properly selecting the subjects in the sample.

Bias and Bias Correction of $\hat{\beta}_1$

It is well known that the least-squares estimates of the $\beta's$ are unbiased if no restrictions are placed on the observations. Since the only $y'_{ij}s$ included in this study are those with $y_{ij} > log(10)$, we have

$$\begin{aligned} E(y_{ij}|y_{ij} > log(10)) &= \beta_0 + \beta_1 t_{ij} + \tau_i + E[(y_{ij} - \beta_0 - \beta_1 t_{ij} - \tau_i)|y_{ij} > log(10)] \\ &= \beta_0 + \beta_1 t_{ij} + \tau_i + E[Z_{ij}|Z_{ij} > z_{ij}], \end{aligned}$$

where Z_{ij} is a normal random variable with mean 0 and $z_{ij} = log(10) - \beta_0 - \beta_1 t_{ij} - \tau_i$. Hence, $E[y_{ij}|y_{ij} > log(10)] = \beta_0 + \beta_1 t_{ij} + \tau_i$ only if $z_{ij} = -\infty$. Utilizing the conditional moments of the random vector $(Z_{i1}, Z_{i2}, \ldots, Z_{ik})$ with correlated components (see Tallis(1961), and McGill(1992)), it can be shown that

$$E(\hat{\beta}_1) = \beta_1 + \frac{n}{[D - n^2 s^2]} \sum_{i=1}^{n} \boldsymbol{a}_i \boldsymbol{\phi}_i,$$

where $\boldsymbol{\phi}_i = E[(Z_{i1}, Z_{i2}, \ldots, Z_{ik})^t | Z_{i1} > z_{i1}, Z_{i2} > z_{i2}, \ldots, Z_{ik} > z_{ik}]$. Thus, $\hat{\beta}_1$ is a biased estimate of β_1. It has been shown that when one conditions the sample so that the $y'_{ij}s$ lie above a tilted line, $\hat{\beta}_1$ becomes unbiased. However, this process results in the loss of additional subjects whose dioxin measurements fall below the tilted line. The process of bias correction is outlined below for $k = 3, 4, 5$.

Special Cases

In the present section, we derive expressions for $\hat{\beta}_1$ for $k = 3, 4, 5$ and obtain conditions under which it is unbiased. For these values of k, we also derive expressions for their conditional variances which are computed and compared to evaluate the benefit or loss accrued from additional measurements per subject.

In what follows, we assume an AR(1) structure for the covariance matrix $\boldsymbol{\Phi}$, and take the successive measurement times Δ units apart, that is, $t_{ij} = t_{i1} + (j-1)\Delta$.

For convenience, we denote $t_{i1} = t_i$ for $i = 1, 2, \ldots, n$.

AR(1) Correlation Matrix, k=3

With the above assumptions,

$$\boldsymbol{\Phi} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix},$$

and it can be seen that

$$\alpha = \frac{3-\rho}{1+\rho}, \qquad ns^2 = \alpha \sum_{i=1}^{n} (t_i - \bar{t})^2,$$

$$CM = \alpha \bar{t}^2 + 2\alpha \bar{t}\Delta + \frac{5 - 4\rho + \rho^2}{1+\rho}\Delta^2,$$

$$ns^2 = \alpha^2 \sum_{i=1}^{n} (t_i - \bar{t})^2, \qquad \bar{t}^* = \alpha(\bar{t} + \Delta),$$

and

$$\boldsymbol{a}^t = \frac{\alpha\Delta}{1-\rho^2}[-1, 0, 1],$$

with $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$. This gives $D - n^2 s^2 = \frac{2\alpha}{(1-\rho)^2} n^2 \Delta^2$ and, consequently,

$$\hat{\beta}_1 = \frac{1}{2n\Delta} \sum_{i=1}^{n} (y_{i3} - y_{i1}).$$

It has been shown that (see Michalek, et. al. (1998)) $\hat{\beta}_1$ is an unbiased estimate of β_1 if, for each subject, $y_{ij} > c_j$ with $c_j = log(10) + (3-j)\Delta\bar{t}$ for $j = 1, 2, 3$. Let $Var_c(X)$ denote the conditional variance of X and $Cov_c(X_1, X_2)$ denote the conditional covariance between X_1, and X_2, under the above conditions. Then

$$Var_c(\hat{\beta}_1) = \frac{1}{4n\Delta^2}[Var_c(y_{i3}) + Var_c(y_{i1}) - 2Cov_c(y_{i3}, y_{i1})].$$

Since, the $\boldsymbol{y}_i$'s are independent and identically distributed, we will denote $\boldsymbol{y}_i$ by $\boldsymbol{Y}^t = (Y_1, Y_2, Y_3)$ for the purpose of computing the expectations, variances and covariances.

It can be seen that under the AR(1) structure, if (Y_1, Y_2, Y_3) has a trivariate normal distribution, the conditional distributions of Y_1 and Y_3 given Y_2 are independent. That is,

$$f(y_1, y_2, y_3) = f_1(y_1|y_2)f_3(y_3|y_2)f_2(y_2),$$

where $f(y_1, y_2, y_3)$ is the joint probability density function (pdf) of (Y_1, Y_2, Y_3), $f_i(y_i|y_j)$ is the conditional pdf of $Y_i|Y_j = y_j$ for $i = 1, 3$ and $j = 2$, and $f_2(y_2)$ is the marginal pdf of Y_2. This observation facilitates the computation of $Var_c(\hat{\beta}_1)$ in terms of univariate integrals, as seen below. Let $S_c = P(Y_1 > c_1, Y_2 > c_2, Y_3 > c_3)$ denote the trivariate joint survival function of Y_1, Y_2, Y_3. Let $\phi(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$. For $i = 1, 2, 3$, let S_i denote the univariate survival functions of Y_i, and

$$a_i(u) = \frac{c_i - \mu_i - \rho\sigma u}{\sigma\sqrt{(1-\rho^2)}}, \tag{4}$$

$$B_i(u) = \mu_i + \rho\sigma u, \tag{5}$$

$$b_i(u) = \mu_i + \rho\sigma u - E_c(Y_i), \tag{6}$$

$$D_i(u) = \mu_i + \sigma u. \tag{7}$$

Let $A = \{u : u > \frac{c_2 - \mu_2}{\sigma}\}$, then

$$S_c = \int_A S_1(a_1(u))S_3(a_3(u))\phi(u)du$$

and

$$E_c(Y_i) = \frac{1}{S_c}\int_A S_j(a_j(u))[B_i(u)S_i(a_i(u)) + \sigma\sqrt{(1-\rho^2)}\phi(a_i(u))]\phi(u)du. \tag{8}$$

Expression (8) gives $E_c(Y_1)$ for $i = 1, j = 3$, and $E_c(Y_3)$ for $i = 3, j = 1$. Let

$$\begin{aligned} Var_c(Y_i) = \frac{1}{S_c}\int_A \; & S_j(a_j(u))[(b_i^2(u)S_i(a_i(u)) + \\ & \sigma^2(1-\rho^2)\{a_i(u)\phi(a_i(u)) + S_i(a_i(u))\} \\ & +2\sigma\sqrt{(1-\rho^2)}b_i(u)\phi(a_i(u))]\phi(u)du. \end{aligned} \tag{9}$$

Then (9) gives $Var_c(Y_1)$ for $i = 1, j = 3$, and $Var_c(Y_3)$ for $i = 3, j = 1$ respectively. Also

$$Cov_c(Y_1, Y_3) = \frac{1}{S_c}\int_A [b_1(u)S_1(a_1(u)) + \sigma\sqrt{(1-\rho^2)}\phi(a_1(u))]$$
$$[b_3(u)S_3(a_3(u)) + \sigma\sqrt{(1-\rho^2)}\phi(a_3(u)]\phi(u)du.$$

<u>AR(1) Correlation Matrix, k=4</u>

For $k = 4$,

$$\mathbf{\Phi} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

,

$$\alpha = \frac{2(2-\rho)}{1+\rho}, \qquad \bar{t}^* = \alpha\bar{t} + \Delta\frac{3(2-\rho)}{1+\rho},$$

$$CM = \alpha\bar{t}^2 + \frac{6(2-\alpha)}{1+\rho}\bar{t} + \frac{5(1-\rho^2) - 6\rho + 9}{1-\rho^2}\Delta^2,$$

and

$$\boldsymbol{a}_i^t = \frac{(2-\rho)}{(1+\rho^2)(1-\rho)}\Delta[-(3-\rho), -(1-\rho^2), (1-\rho^2), (3-\rho)].$$

This gives

$$\hat{\beta}_1 = \frac{1}{(10 - 5\rho + \rho^2)(n\Delta)}\sum_{i=1}^{n}[(3-\rho)(y_{i4} - y_{i1}) + (1-\rho^2)(y_{i3} - y_{i2})].$$

As in the case of $k = 3$, $\hat{\beta}_1$ can be made unbiased by selecting $c_i = log(10) + (4-i)\Delta\bar{t}$, and by restricting $Y_i > c_i$, for $i = 1, \ldots, 4$. Let Var_c and Cov_c denote the conditional variance and covariance of the components of $(y_{i1}, y_{i2}, y_{i3}, y_{i4})$, then

$$\begin{aligned} Var_c(\hat{\beta}_1) &= \frac{1}{n\Delta^2(10-5\rho+\rho^2)}[(3-\rho)^2(Var_c(y_{i1}) + Var_c(y_{i4}) - 2Cov_c(y_{i1}, y_{i4})) \\ &+ (1-\rho^2)^2(Var_c(y_{i2}) + Var_c(y_{i3}) - 2Cov_c(y_{i2}, y_{i3})) \\ &+ 2(3-\rho)(1-\rho^2)(Cov_c(y_{i1}, y_{i2}) - Cov_c(y_{i1}, y_{i3}) \\ &+ Cov_c(y_{i3}, y_{i4}) - Cov_c(y_{i2}, y_{i4})]. \end{aligned} \tag{10}$$

Again denoting $\boldsymbol{y_i}$ by $\boldsymbol{Y}^t = (Y_1, Y_2, Y_3, Y_4)$ the conditional means, variances and covariances in (10) can be expressed in terms of bivariate integrals. Under the AR(1) structure, it can be seen that

$$\begin{bmatrix} Y_1 \\ Y_4 \end{bmatrix} \Bigg| \begin{bmatrix} Y_2 = y_2 \\ Y_3 = y_3 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 + \rho(y_2 - \mu_2) \\ \mu_4 + \rho(y_3 - \mu_3) \end{bmatrix}, \sigma^2 \begin{bmatrix} 1-\rho^2 & 0 \\ 0 & 1-\rho^2 \end{bmatrix} \right).$$

Let $f(u,v)$ denote the joint pdf of a bivariate normal distribution given by

$$f(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}(u^2 - 2\rho uv + v^2)}.$$

For $i = 1,2,3,4$, let $a_i(u)$, $B_i(u)$, $b_i(u)$, $D_i(u)$ be as in equations (4)-(7), and

$$d_i(u) = \mu_i + \sigma u - E_c(Y_i). \tag{11}$$

Let $A = \{u : u > \frac{c_2 - \mu_2}{\sigma}\}$ and $B = \{v : v > \frac{c_3 - \mu_3}{\sigma}\}$. Then

$$S_c = \int_A \int_B S_1(a_1(u)) S_4(a_4(v)) f(u,v) du dv,$$

$$E_c(Y_1) = \frac{1}{S_c} \int_A \int_B S_4(a_4(v))[B_1(u)S_1(a_1(u)) + \sigma\sqrt{(1-\rho^2)}\phi(a_1(u))] f(u,v) du dv,$$

$$E_c(Y_4) = \frac{1}{S_c} \int_A \int_B S_1(a_1(u))[B_4(v)S_4(a_4(v)) + \sigma\sqrt{(1-\rho^2)}\phi(a_4(v))] f(u,v) du dv,$$

$$E_c(Y_2) = \frac{1}{S_c} \int_A \int_B D_2(u) S_1(a_1(u)) S_4(a_4(v)) f(u,v) du dv,$$

$$E_c(Y_3) = \frac{1}{S_c} \int_A \int_B D_3(v) S_1(a_1(u)) S_4(a_4(v)) f(u,v) du dv,$$

$$\begin{aligned} Var_c(Y_1) = \frac{1}{S_c} \int_A \int_B \; & S_4(a_4(v))[b_1^2(u) S_1(a_1(u)) + \sigma^2(1-\rho^2) \\ & \{a_1(u)\phi(a_1(u)) + S_1(a_1(u))\} + 2\sigma\sqrt{(1-\rho^2)} \\ & b_1(u)\phi(a_1(u))] f(u,v) du dv, \end{aligned}$$

$$\begin{aligned} Var_c(Y_4) = \frac{1}{S_c} \int_A \int_B \; & S_1(a_1(u))[b_4^2(v) S_4(a_4(v)) + \sigma^2(1-\rho^2) \\ & \{a_4(v)\phi(a_4(v)) + S_4(a_4(v))\} + 2\sigma\sqrt{(1-\rho^2)} \\ & b_4(v)\phi(a_4(v))] f(u,v) du dv, \end{aligned}$$

$$Var_c(Y_2) = \frac{1}{S_c}\int_A\int_B d_2^2(u)S_1(a_1(u))S_4(a_4(v))f(u,v)dudv. \tag{12}$$

$Var_c(Y_3)$ can be obtained from (12) by replacing $d_2^2(u)$ with $d_3^2(v)$. For the covariances we have,

$$Cov_c(Y_1,Y_2) = \frac{1}{S_c}\int_A\int_B \, d_2(u)[b_1(u)S_1(a_1(u)) + \sigma\sqrt{(1-\rho^2)}\phi(a_1(u))] \\ S_4(a_4(v))f(u,v)dudv. \tag{13}$$

$Cov_c(Y_1,Y_3)$ is obtained from (13) by replacing $d_1(u)$ with $d_3(v)$. We also have

$$Cov_c(Y_1,Y_4) = \frac{1}{S_c}\int_A\int_B \, [b_1(u)S_1(a_1(u)) + \sigma\sqrt{(1-\rho^2)}\phi(a_1(u))] \\ [b_4(v)S_4(a_4(v)) + \sigma\sqrt{(1-\rho^2)}\phi(a_4(v))]f(u,v)dudv,$$

$$Cov_c(Y_2,Y_4) = \frac{1}{S_c}\int_A\int_B \, d_2(u)[b_4(v)S_4(a_4(v)) + \sigma\sqrt{(1-\rho^2)}\phi(a_4(v))] \\ S_1(a_1(u))f(u,v)dudv. \tag{14}$$

$Cov_c(Y_3,Y_4)$ is obtained from (14) by replacing $d_2(u)$ by $d_3(v)$.

<u>AR(1) Correlation Matrix, k=5</u>

For $k = 5$,

$$\mathbf{\Phi} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}.$$

It can be seen that in this case,

$$\alpha = \frac{5-3\rho}{1+\rho}.$$

For this case

$$\hat{\beta}_1 = \frac{1}{(2n\Delta)(\rho^2-4\rho+5)}\sum_{i=1}^{n}[(2-\rho)(y_{i5}-y_{i1}) + (1-\rho)^2(y_{i4}-y_{i2})].$$

As in the case $k = 3$, $\hat{\beta}_1$ can be made unbiased by selecting $c_i = log(10)+(5-i)\Delta\bar{t}$, and by restricting $Y_i > c_i$, for $i = 1, 2, 3, 5$. Let Var_c and Cov_c denote the variance and covariance of the components of $(y_{i1}, y_{i2}, y_{i2}, y_{i4}, y_{i5})$ under the above condition, then

$$\begin{aligned} Var_c(\hat{\beta}_1) &= \frac{1}{4n\Delta^2(\rho^2 - 4\rho + 5)}[(2-\rho)^2(Var_c(y_{i1}) + Var_c(y_{i5}) - 2Cov_c(y_{i1}, y_{i5})) \\ &+ (1-\rho)^4(Var_c(y_{i2}) + Var_c(y_{i4}) - 2Cov_c(y_{i2}, y_{i4})) \\ &+ 2(2-\rho)(1-\rho)^2(Cov_c(y_{i1}, y_{i2}) - Cov_c(y_{i1}, y_{i4}) \\ &+ Cov_c(y_{i4}, y_{i5}) - Cov_c(y_{i2}, y_{i5})]. \end{aligned} \tag{15}$$

The conditional means, variances and covariances in (15) can be expressed in terms of trivariate integrals, as seen below.

Under the AR(1) structure, it can be seen that

$$\left[\begin{array}{c} Y_1 \\ Y_5 \end{array}\right] \Bigg\| \left[\begin{array}{c} Y_2 = y_2 \\ Y_3 = y_3 \\ Y_4 = y_4 \end{array}\right] \sim N\left(\left[\begin{array}{c} \mu_1 + \rho(y_2 - \mu_2) \\ \mu_5 + \rho(y_4 - \mu_4) \end{array}\right], \sigma^2 \left[\begin{array}{cc} 1-\rho^2 & 0 \\ 0 & 1-\rho^2 \end{array}\right]\right).$$

Let $f(u, v, w)$ denote the joint pdf of a trivariate normal distribution, given by

$$f(u, v, w) = \frac{1}{(2\pi)^{3/2}(1-\rho^2)} e^{-\frac{1}{2}(u^2 - 2\rho(uv+vw)+(1+\rho^2)v^2+w^2)}.$$

Let $a_i(u)$, $b_i(u)$, $B_i(u)$, $D_i(u)$, and $d_i(u)$,$i = 1, 2, \ldots, 5$ be as defined in (4)-(7), and (11). In addition to the sets A, and B, let $C = \{w : w > \frac{c_4 - \mu_4}{\sigma}\}$. Then, the 5-dimensional survival function is given by

$$S_c = \int_A \int_B \int_C S({}_1a_1(u))S_5(a_5(w))f(u, v, w)dudvdw.$$

The expected values, variances and covariances are:

$$E_c(Y_1) = \frac{1}{S_c}\int_A \int_B \int_C [B_1(u)S_1(a_1(u)) + \sigma\sqrt{(1-\rho^2)}\phi(a_1(u)]S_5(a_5(w))f(u, v, w)dudvdw,$$

$$E_c(Y_5) = \frac{1}{S_c}\int_A \int_B \int_C [B_5(w)S_5(a_5(w)) + \sigma\sqrt{(1-\rho^2)}\phi(a_5(w)]S_1(a_1(u))f(u, v, w)dudvdw,$$

$$E_c(Y_2) = \frac{1}{S_c}\int_A\int_B\int_C D_2(u)S_1(a_1(u))S_5(a_5(w))f(u,v,w)dudvdw,$$

$$E_c(Y_4) = \frac{1}{S_c}\int_A\int_B\int_C D_4(w)S_1(a_1(u)S_5(a_5(w))f(u,v,w)dudvdw,$$

$$\begin{aligned} Var_c(Y_1) = \frac{1}{S_c}\int_A\int_B\int_C \; & S_5(a_5(w))[b_1^2(u)S_1(a_1(u)) + +2\sigma\sqrt{(1-\rho^2)}b_1(u)\phi(a_1(u)) + \\ & \sigma^2(1-\rho^2)\{a_1(u)\phi(a_1(u)) + S_1(a_1(u))]f(u,v,w)dudvdw, \end{aligned}$$

$$\begin{aligned} Var_c(Y_5) = \frac{1}{S_c}\int_A\int_B\int_C \; & S_1(a_1(u))[b_5^2(w)S_5(a_5(w)) + 2\sigma\sqrt{(1-\rho^2)}b_5(w)\phi(a_5(w)) + \\ & \sigma^2(1-\rho^2)\{a_5(w)\phi(a_5(w)) + S_5(a_5(w))]f(u,v,w)dudvdw, \end{aligned}$$

$$Var_c(Y_2) = \frac{1}{S_c}\int_A\int_B\int_C d_2^2(u)S_1(a_1(u))S_5(a_5(w))f(u,v,w)dudvdw.$$

and

$$Var_c(Y_4) = \frac{1}{S_c}\int_A\int_B\int_C d_4^2(w)S_1(a_1(u))S_5(a_5(w))f(u,v,w)dudvdw.$$

The covariances are:

$$\begin{aligned} Cov_c(Y_1, Y_2) = \frac{1}{S_c}\int_A\int_B\int_C \; & d_2(u)[b_1(u)S_1(a_1(u)) + \sigma\sqrt{(1-\rho^2)}\phi(a_1(u))] \\ & S_5(a_5(w))f(u,v,w)dudvdw, \end{aligned}$$

$$\begin{aligned} Cov_c(Y_1, Y_4) = \frac{1}{S_c}\int_A\int_B\int_C \; & d_4(w)[b_1(u)S_1(a_1(u)) + \sigma\sqrt{(1-\rho^2)}\phi(a_1(u))] \\ & S_5(a_5(w))f(u,v,w)dudvdw, \end{aligned}$$

$$\begin{aligned} Cov_c(Y_1, Y_5) = \frac{1}{S_c}\int_A\int_B\int_C \; & [b_1(u)S_1(a_1(u)) + \sigma\sqrt{(1-\rho^2)}\phi(a_1(u))] \\ & [b_5(w)S_5(a_5(w)) + \sigma\sqrt{(1-\rho^2)}\phi(a_5(w))] \\ & f(u,v,w)dudvdw, \end{aligned}$$

$$Cov_c(Y_2, Y_4) = \frac{1}{S_c} \int_A \int_B \int_C d_2(u) d_4(w) S_1(a_1(u)) S_5(a_5(w) f(u, v, w) du dv dw,$$

$$Cov_c(Y_2, Y_5) = \frac{1}{S_c} \int_A \int_B \int_C d_2(u)[b_5(w) S_5(a_5(w)) + \sigma\sqrt{(1-\rho^2)}\phi(a_5(w))] \\ S_1(a_1(u)) f(u, v, w) du dv dw,$$

and

$$Cov_c(Y_4, Y_5) = \frac{1}{S_c} \int_A \int_B \int_C d_4(w)[b_5(w) S_5(a_5(w)) + \sigma\sqrt{(1-\rho^2)}\phi(a_5(w))] \\ S_1(a_1(u)) f(u, v, w) du dv dw.$$

Results and Discussion

The following table gives the standard error of $\hat{\beta}_1$ after bias correction, for $k = 3, 4, 5$. It also gives 95% confidence intervals for β_1 and the half-life, $t_{1/2}$, along with their widths.

k	std error $\hat{\beta}_1$	$95\% CI$ β_1	Width β_1	$95\% CI$ $t_{1/2}$	Width $t_{1/2}$
3	.004089	(.0732, .0893)	.01603	(7.765, 9.465)	1.6999
4	.002729	(.0759, .0866)	.01069	(8.004, 9.133)	1.1282
5	.002863	(.0756, .0869)	.01123	(7.980, 9.164)	1.1841

From the above table, we can assess the effect of the loss of subjects due to the selection for correcting the bias in the least-squares estimate of β_1 on the standard error of $\hat{\beta}_1$, widths of the 95% confidence intervals for β_1 and $t_{1/2}$. There is a 37.2% reduction in the standard error of $\hat{\beta}_1$ form $k = 3$ to $k = 4$, while there is a 4.69% increase from $k = 4$ to $k = 5$. In the 95% confidence interval for β_1, there is a 1.15% reduction from $k = 3$ to $k = 4$, while there is a 4.73% increase from $k = 4$ to $k = 5$. Similarly, in the width of the 95% confidence interval for $t_{1/2}$, there is a 33.26% reduction from $k = 3$ to $k = 4$, while there is a 4.9% increase from $k = 4$ to

$k = 5$. Thus, there is a loss in precision of the estimates resulting from the loss of subjects due to truncation. This comparison can be more dramatic if larger values of k are considered for these comparisons.

References

1.IMSL, Inc. (1991), IMSL Math/Library Users Manual, Version 2.0, IMSL, Houston.

2.McGill, J. I. (1992). The multivariate hazard gradient and moments of the truncated multinormal distribution. *Communications in Statistics, Theory and Methods* **21**, 3053 – 3060.

3.Michalek, J. M., Tripathi, R. C., Kulkarni, P., Selvavel, K., and Gupta, P. L. (1998). Correction for bias introduced by truncation in pharmacokinetic studies of environmental contaminants. *Environmetrics* **9**, 165 – 174.

4.Numerical Algorithms Group (1995). Fortran Library Mark 17. NAG Limited, Oxford.

5.Tallis, G. M. (1961). The moment generating function of the truncated multinormal distribution. *Journal of the Royal Statistical Society* **B, 23**, 223 – 229.

6.Tripathi, Ram C. (1998) Regression to the mean in half-life studies. Air Force Research Laboratory, Brooks AFB, Texas, *AFRL-HE-BR-TR-1998-0001.*

```
      parameter(nmax=10)
      implicit real*8(a-h,o-z)
      external dqand,fxyz,dnordf
      dimension cut(5),xmu(5),sig(nmax,nmax)
     @,bb(3),ey(16),k(15),aa(3)
c To compute variance of UWLS bias corrected beta1 based on truncated data k=5
c *****variance computed using E(y-mu)**2 rather than E(y**2)-mu**2 *******
      common /a/twopi,rt2pi,rho,sigma,rtrho,xlim1,ylim5
     @,c1,c2,c3,c4,c5,ul
      common/c/kk
      common /b/ cut,xmu,ey
      open(unit=8,name='var5bt1.out')
      twopi=6.283185307d0
      rt2pi=dsqrt(twopi)
      rho=.8557d0
      sigma=.8d0
      beta0=5.37850503d0
      tbar=14.8368d0
      beta1=-0.08124682d0
      delta=5.0d0
      cut(5)=dlog(10.0d0)
      cut(4)=cut(5)-delta*beta1
      cut(3)=cut(5)-2.0d0*delta*beta1
      cut(2)=cut(5)-3.0d0*delta*beta1
      cut(1)=cut(5)-4.0d0*delta*beta1
      xmu(1)=beta0+beta1*tbar
      xmu(2)=beta0+beta1*(tbar+delta)
      xmu(3)=beta0+beta1*(tbar+2.d0*delta)
      xmu(4)=beta0+beta1*(tbar+3.d0*delta)
      xmu(5)=beta0+beta1*(tbar+4.d0*delta)
      rtrho=sigma*dsqrt(1.d0-rho*rho)
      n=213
      cut5=(cut(5)-xmu(5))/sigma
      tail5=1.d0-dnordf(cut5)
      write(8,350)
 350  format('****Variance Computation of beta1 with Five Measurements*
     @ ****'//)
      write(8,*)'Probability of y5 > log(10)= ',tail5
      n5=idint(tail5*dfloat(n))
      write(8,*)'For 92,n=',n,',  For 98,predicted n=',n5
      write(8,*)
      write(8,*)'rho=',rho,' Sigma=',sigma
      do 10 i=1,5
      do 10 j=1,5
        if(i .eq. j)then
        sig(i,j)=sigma**2
        else
        sig(i,j)=0.0d0
```

```
        sig(i,j)=sigma**2*rho**(abs(i-j))
        end if
 10     continue
      do i=1,15
       k(i)=i
      end do
       write(8,*)'xmu s are:',(xmu(i),i=1,5)
       mincls=1000
c To check if the total probability sc integrates to 1
       kk=16
       c1=-2.0d0
       c2=-2.0d0
       c3=-2.0d0
       c4=-2.0d0
       c5=-2.0d0
       ul=10.00d0
       ya=(c2-xmu(2))/sigma
       yb=(ul-xmu(2))/sigma
       aa(1)=ya
       aa(2)=(c3-xmu(3))/sigma
       aa(3)=(c4-xmu(4))/sigma
       bb(1)=yb
       bb(2)=yb
       bb(3)=yb
       maxfcn=16777216
       errrel=.001d0
       errabs=.001d0
       call dqand(fxyz,3,aa,bb,errabs,errrel,maxfcn,ey(kk),errest)
      write(6,111)aa(1),aa(2),aa(3),bb(1),bb(2),bb(3),ey(16)
 111  format(' Lower limits=',3(f10.6,2x),/
     @' Upper limits= ',3(f10.6,2x), 'Total Prob= ',f18.12)
      write(8,111)aa(1),aa(2),aa(3),bb(1),bb(2),bb(3),ey(16)
      kk=15
      minpts=0
      ya=(cut(2)-xmu(2))/sigma
      yb=(ul-xmu(2))/sigma
      aa(1)=ya
      aa(2)=(cut(3)-xmu(3))/sigma
      aa(3)=(cut(4)-xmu(4))/sigma
      bb(1)=yb
      bb(2)=yb
      bb(3)=yb
      write(8,*)'beta0',beta0,' Beta1=',Beta1
      call dqand(fxyz,3,aa,bb,errabs,errrel,maxfcn,ey(kk),errest)
      sc=ey(kk)
      write(6,112)aa(1),aa(2),aa(3),bb(1),bb(2),bb(3),ey(15)
 11   format(' Lower limits=',3(f10.6,2x),/
     @' Upper limits= ',3(f10.6,2x), 'Upp Tail Sc= ',f18.12)
      write(8,112)aa(1),aa(2),aa(3),bb(1),bb(2),bb(3),ey(15)
```

```
        do i=1,14
        kk=k(i)
      ya=(cut(2)-xmu(2))/sigma
      yb=(ul-xmu(2))/sigma
      aa(1)=ya
      aa(2)=(cut(3)-xmu(3))/sigma
      aa(3)=(cut(4)-xmu(4))/sigma
      bb(1)=yb
      bb(2)=yb
      bb(3)=yb
      call dqand(fxyz,3,aa,bb,errabs,errrel,maxfcn,ey(kk),errest)
      ey(kk)=ey(kk)/sc
      write(6,*)'ey are Kk=',Kk,'  ey(Kk):',ey(kK)
      write(8,*)'ey are Kk=',Kk,'  ey(Kk):',ey(kK)
      end do
      write(8,*)'expected values are:',(ey(kk),kk=1,15)
      write(6,*)'expected values are:',(ey(kk),kk=1,15)
c to compute variances and covariances
      sc=ey(15)
      s11=ey(5)
      s22=ey(6)
      s44=ey(7)
      s55=ey(8)
      s12=ey(9)
      s14=ey(10)
      s15=ey(11)
      s24=ey(12)
      s25=ey(13)
      s45=ey(14)
      den=dfloat(4*n5)*(delta**2)*(5.d0-4.d0*rho+rho**2)**2
      write(6,*)'den=',den
      trm1=(rho-2.d0)**2*(s11+s55-2.d0*s15)
      trm2=((1.d0-rho)**4)*(s44+s22-2.*s24)
      trm3=2.*(rho-2.d0)*((1.-rho)**2)*(s14-s12-s45+s25)
      write(8,*)'terms are:',trm1,trm2,trm3
      varbt1=(trm1+trm2+trm3)/den
      sdbt1=dsqrt(varbt1)
      write(8,*)'varbt1=',varbt1,' sdbt1=',sdbt1
      hl=dlog(2.d0)/(-beta1)
      xllm95=-beta1-1.96d0*sdbt1
      xulm95=-beta1+1.96d0*sdbt1
      wlm=xulm95-xllm95
      hll95=dlog(2.d0)/xulm95
      hlu95=dlog(2.d0)/xllm95
      whl=hlu95-hll95
      write(8,* )'For the Decay Rate:'
      write(8,100)-beta1,xllm95,xulm95,wlm
  100 format(' Decay rate= ',f9.6/' 95% CI: (',f10.6,' , ',f10.6,' )
     @ '/ 'Width=', f10.6)
```

```
      write(8,*)'For the Half Life:'
      write(8,200)hl,hl195,hlu95,whl
 200  format(' Half Life= ',f10.5/' 95% CI: (',f10.6,' , ',f10.6,')
     @ '/ ' Width= ',f10.6)
      stop
      end
c Subroutine fxyz to calculate the integrand
      function fxyz(n,zz)
      implicit real*8(a-h,o-z)
      external dnordf
      dimension xmu(5),cut(5),ey(16),zz(n)
      common /a/twopi,rt2pi,rho,sigma,rtrho,xlim1,ylim5
     @,c1,c2,c3,c4,c5,ul
      common/c/kk
      common /b/cut,xmu,ey
      x=zz(1)
      y=zz(2)
      z=zz(3)
      if(kk .eq. 6) then
      end if
c  x=y2, y=y3, z=y4
      cmean1=xmu(1)+rho*sigma*x
      cmean5=xmu(5)+rho*sigma*z
      probx1=1.d0-dnordf(xlim1)
      proby5=1.d0-dnordf(ylim5)
      phi1=dexp(-.5d0*xlim1**2)/rt2pi
      phi4=dexp(-.5d0*ylim5**2)/rt2pi
      const=(1.-rho*rho)
      qxyz=dexp(-.50d0*(x**2-2.d0*rho*(x*y+y*z)+(1.+rho**2)*y**2+
     @z**2)/const)
      qxyz=qxyz/((twopi**1.5)*const)
      if(kk .eq. 16) then
      xlim=(c1-xmu(1)-rho*sigma*x)/rtrho
      ylim=(c5-xmu(5)-rho*sigma*z)/rtrho
      probx=1.d0-dnordf(xlim)
      proby=1.d0-dnordf(ylim)
      fxyz=probx*proby*qxyz
       else if(kK .eq. 15) then
c compute integrand for Sc
      fxyz=probx1*proby5*qxyz
       else if( kk.eq.1) then
c   compute integrand for E(y1)
       fxyz=proby5*(cmean1*probx1+rtrho*phi1)*qxyz
       else if(kk .eq. 4)then
c computes integrand for E(y5)
       fxyz=probx1*(cmean5*proby5+rtrho*phi5)*qxyz
       else if(kk .eq.2) then
c computes integrand for E(y2)
       fxyz=(xmu(2)+sigma*x)*probx1*proby5*qxyz
```

```
        else if (kk .eq. 3) then
c computes integrand for E(y4)
        fxyz=(xmu(4)+sigma*z)*probx1*proby5*qxyz
        else if (kk .eq. 9)then
c computes E(y1y2)
        fxyz=(xmu(2)+sigma*x-ey(2))*((cmean1-ey(1))*probx1+rtrho*phi1)
     @ *proby5*qxyz
        else if(kk .eq. 10) then
c computes E(y1y4)
        fxyz=((cmean1-ey(1))*probx1+rtrho*phi1)*(xmu(4)+sigma*z-ey(3))
     @ *proby5*qxyz
        else if(kk .eq. 11) then
c computes E(y1y5)
        fxyz=((cmean1-ey(1))*probx1+rtrho*phi1)*((cmean5-ey(4))*proby5
     @ +rtrho*phi5)*qxyz
        else if(kk .eq. 12) then
c  computes integrand for E(y2y4)
        fxyz=(xmu(2)+sigma*x-ey(2))*probx1*(xmu(4)+sigma*z-ey(3))
     @ *proby5*qxyz
        else if(kk .eq. 13) then
c computes integrand for E(y2y5)
        fxyz=(xmu(2)+sigma*x-ey(2))*probx1*((cmean5-ey(4))*proby5+rtrho
     @ *phi5)*qxyz
        else if(kk .eq. 14) then
c computes integrand for E(y4y5)
        fxyz=probx1*(xmu(4)+sigma*z-ey(3))*((cmean5-ey(4))*proby5+rtrho
     @ *phi5)*qxyz
        else if(kk .eq. 5) then
c computes the integrand for E(y1**2)
        fxyz=((cmean1-ey(1))**2*probx1+rtrho**2*(xlim1*phi1+probx1)
     @ +2.*rtrho*(cmean1-ey(1))*phi1)*proby5*qxyz
        else if(kk .eq. 6) then
c computes integrand for E(y2**2)
        fxyz=(xmu(2)+sigma*x-ey(2))**2*probx1*proby*qxyz
        else if(kk .eq. 7) then
c computes integrand for E(y4**2)
        fxyz=(xmu(4)+sigma*z-ey(3))**2*probx1*proby5*qxyz
        else if(kk .eq. 8) then
c computes E(y5**2)
        fxyz=((cmean5-ey(4))**2*proby5+rtrho**2*(limy5*phi5+proby5)
     @ +2.*rtrho*(cmean5-ey(4))*phi5)*probx1*qxyz
        end if
c        write(6,*)'from fxy function, fxy=',fxy
       return
        end
```

DETECTING BIDIMENSIONALITY IN RESPONSE DATA: AN EMPIRICAL TASK ANALYSIS TECHNIQUE

Edward W. Wolfe
Assistant Professor
Foundations of Education

University of Florida
1403 Norman Hall
Gainesville, FL 32611

Final Report for:
Summer Faculty Research Program
Armstrong Research Site

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

And

Armstrong Research Site

July 1998

DETECTING BIDIMENSIONALITY IN RESPONSE DATA: AN EMPIRICAL TASK ANALYSIS TECHNIQUE

Edward W. Wolfe
Assistant Professor
Foundations of Education
University of Florida

Abstract

There are several common situations in which we might expect test data to exhibit multidimensionality across measurement occasions or between subgroups, and task analysis is often used to uncover the multidimensionality inherent in a set of test items or other cognitive tasks. The purpose of this paper is to describe a method for utilizing knowledge of group differences in task performance for the purpose of performing empirical task analysis. To this end, two-dimensional, dichotomous response data were simulated to portray several situations that might arise in operational settings. The factors manipulated during data generation included (a) the magnitude of the relationship between the two dimensions, (b) the proportion of tasks belonging to each dimension and (c) the number of respondents upon whom the tasks were calibrated. Task dimensionality was recovered using a Rasch standardized fit statistic to measure the difference between task calibrations that were estimated on two groups of simulated examinees. Overall, the technique was successful in recovering task dimensionality with 85% accuracy.

DETECTING BIDIMENSIONALITY IN RESPONSE DATA: AN EMPIRICAL TASK ANALYSIS TECHNIQUE

Edward W. Wolfe

Introduction

There are several situations in which test data (or any other cognitive task) might exhibit multidimensionality across measurement occasions or between subgroups. For example, investigations of differential item functioning (DIF) attempt to determine whether a test item measures different constructs in subgroups. Alternatively, trainees who participate in different educational or training programs might think about a particular content domain different ways after being trained. As a result of their experiences, these trainees might be expected to increase not only the amount that they know about the domain but also the way that they structure their knowledge in that domain. As a third example, the knowledge representations of experts and novices in many fields differ in several ways. In operational settings, it is often useful to identify these differences for the purpose of hierarchically depicting the accumulation of expert-like knowledge in that domain. The purpose of this paper is to describe a method for utilizing the multidimensionality inherent in these data for performing task analyses.

Context

This study arose from an interest in determining whether individuals who participate in different training programs (e.g., lecture versus intelligent tutoring system—ITS) structure their knowledge about the domain in which they are trained differently. Under both types of training, trainees should understand more of the content as a result of the training, but they might also construct their understandings of this content in qualitatively different ways. Such structural differences would manifest themselves on different dimensions of the outcome measures of the training programs. This paper reports the results of an investigation of a technique that is based on Rasch measurement theory that can be used to identify the nature of these cognitive structural differences.

Task analyses are commonly used in research and operational settings so that models can be built to depict the manner in which individuals perform a cognitive task. The purpose of task analysis is to identify the strategies and processes that underlie performance of the task (Kail & Bisanz, 1982). Related methods include (a) the *analysis of response patterns* (in which response patterns of individuals are compared to the cognitive model generated from a task analysis for the purpose of evaluating the cognitive model—Siegler, 1976), (b) *analysis of response latencies* (in which processing models that take into account the speed of processing involved in various cognitive activities are formulated, and latencies in the responses of individuals to tasks in which various task parameters are systematically manipulated are compared to determine which processing model best explains performance on the task—Mumaw, Pellegrino, & Glaser, 1980), and (c) *think aloud verbal reports* (in which the verbal reports of individuals who are instructed to "think aloud" are used as the basis for identifying the knowledge and cognitive procedures that are brought into play when performing a cognitive task–Ericsson & Simon, 1982). Although these methods have proven useful in the study of a variety of cognitive tasks, there are two problems. First, these methods require intensive and time consuming data collection and analysis. Second, they require data analysts to make subjective interpretations and judgments about the individual's thought patterns. Nevertheless, task analytic methods have found use in several settings, including applications to identifying dimension loadings in multidimensional item response theory (IRT) (DiBello, Stout, & Roussos, 1995; Kelderman, 1994; Tatsuoka, 1990).

Empirically based methods derived from mathematical modeling of task or item responses provide a potential alternative to traditional task analytic methods. In the context of these mathematical models, the problem of differentiating between qualitatively different cognitive structures becomes one of capturing the multidimensionality inherent in the responses as manifested by the groups of interest. Interestingly, statistics that are associated with unidimensional IRT models may be useful in a task such as this. Hambleton and Swaminathan (1985) suggest three ways that this can be accomplished in their discussion of methods for

evaluating DIF: (a) comparing item characteristic curves, (b) comparing item parameters, and (c) comparing model-data fit statistics. The technique explored in this study utilizes item parameter comparisons.

Technique

The empirical task analytic technique employed in this study relies on three assumptions—(a) at least partial multidimensionality exists in the response data, (b) task (item) parameter estimates will *not* remain stable across subgroups or measurement conditions as a result of this multidimensionality (i.e., the correlation between the two dimensions is less than perfect), and (c) groups of examinees can be differentiated based on the degree to which they manifest each dimension contained in the tasks. In addition, the technique requires that tasks "define" these multiple dimensions through one set of tasks loading solely on one dimension and the remaining tasks loading on different dimensions depending on the cognitive characteristics of the group responding to those tasks. In the following example, only bidimensional applications are explored. Finally, group membership must be identifiable via an external criterion. If these requirements are reasonable for a particular application, then the following procedure may be used to identify the nature of the multidimensionality inherent in the data.

The procedure follows.

Step 1: Identify individuals who belong to each of two groups (e.g., *lecture* versus *ITS*), The two groups must differ in the construct being measured on at least some of the tasks on which individuals are measured. That is, some subset of the tasks must exhibit bidimensionality with respect to group membership while the remaining tasks must demonstrate unidimensionality across groups. This is not an uncommon situation; an entire line of research has developed around a specific case of this phenomenon—differential item functioning. Once the two groups are identified, the tasks are administered to each individual in each group.

Step 2: Calibrate the response data for each group separately using a unidimensional IRT model. These analyses will yield two separate task calibrations—one based on each group. These task calibrations can be compared using a standardized difference (t_d). This statistic

depicts the degree to which the two task parameter estimates differ by a magnitude that is more than that likely by chance alone. This allows us to identify which tasks define the two dimensions being measured by the instrument. More specifically, we assume that the calibrations for the tasks that define the second dimension will differ between the two groups.

Item Response Models

There are several item response models that can be used to calibrate the tasks, each model requiring only that responses to a task can be portrayed using numerical codes that depict the degree to which a particular response is characteristic of a trait of interest. It is important to note that these codes are *ordinal* in nature. Commonly, these codes represent the degree of correctness (e.g., right/wrong or degrees of partial credit), but the codes could also represent other characteristics such as speed of processing or counts of behaviors. These numerical codes may be any integer, provided the ordinality of the integers accurately portray the degree of the trait being measured. In this paper, only dichotomous data (e.g., 1s and 0s) are explored, but additional models could be employed to portray polytomous, binomial, or Poisson processes (Wright & Masters, 1982).

The dichotomous Rasch model (Equation 1) depicts the probability that a respondent will be rated with a 1 (versus a 0) with the coding system. This probability is portrayed as a function of two parameters—θ_r, the respondent's location on the underlying linear continuum (e.g., ability) and δ_t the task's location on the underlying continuum (e.g., difficulty). When response data are calibrated to the dichotomous Rasch model, a parameter estimate and a standard error for that parameter estimate is generated for each element (i.e., each individual respondent and task)of the measurement context. In the technique described above, a separate calibration is obtained for each task from each of two groups of respondents.

$$P(x|\theta,\delta) = \frac{\exp(\theta_r - \delta_t)}{1 + \exp(\theta_r - \delta_t)} \qquad (1)$$

Once the pair of calibrations has been obtained for each task, the invariance of these calibrations across groups of respondents can be evaluated. To accomplish this, the standardized difference of the two task calibrations is examined (Equation 2). This statistic represents the degree to which differences between pairs of observed task calibrations was likely to have been produced solely by random variation. When the number of tasks is large, t_d has an expected value of zero and variance of approximately one. Tasks for which calibrations from different groups of examinees are largely different are suspected to belong to different dimensions, depending on which group of respondents the calibration is based.

$$t_d = \frac{(\delta_1 - \delta_2)}{\sqrt{SE^2_{\delta_1} + SE^2_{\delta_2}}} \tag{2}$$

Method

To investigate the usefulness of this procedure for uncovering the cognitive characteristics that differentiate groups, a simulation study was performed. This study focused on the decision accuracy of the empirical task analytic technique when several parameters of the measurement context are manipulated. To this end, data were generated by altering three factors: (a) the number of respondents upon whom task calibrations are estimated (N), (b) the strength of the correlation between the two task dimensions (r_{12}), and (c) the proportion of tasks that belong to the second dimension (p_2). Levels for each of these factors were: (a) $N \in (500{,}2000)$, (b) $r_{12} \in (.25, .50, .75)$, and (c) $p_2 \in (.25,.50,.75)$. For each combination of these factors, 30 pairs of replications were generated using common seed values across the various conditions—the pairs of replications depicting two respondent groups. Thus, a total of 1080 data sets were generated (30 replications x 2 types of respondents x 2 respondent sample sizes x 3 levels of correlation between the task dimensions x 3 levels of the proportion of tasks belonging to each dimension).

Each data set was generated using IRT methods described by Harwell, Stone, Hsu, & Krisci (1996). More specifically, the probability of being coded as a "1" was generated for each respondent based on Equation 1. For the first group of respondents, the group designated to be "unidimensional," θ_{1r} was sampled from a N(0,1) distribution. For the second group, designated as "bidimensional," θ_{2r} was sampled to represent a N(0,1) distribution using linear regression. That is, $\theta_{2r}= r_{12}\theta_{1r}+\varepsilon$, where ε was sampled from a $N(0,S_\varepsilon^2)$ distribution and $S_\varepsilon^2=1-r_{12}$. The task parameters for the unidimensional group, δ_{1t}, were sampled from a N(0,1) distribution with absolute values truncated to a limit of 2. This value was used to generate task responses for unidimensional and bidimensional respondents for the unidimensional tasks. It was also used for the bidimensional tasks for the unidimensional respondents. The task parameters for the bidimensional tasks for the bidimensional respondents, δ_{2t}, were transformed using linear regression so that $\delta_{2t}= r_{12}\delta_{1t}+S_\varepsilon^2$ and were truncated so that the absolute values did not exceed 2. In all cases, 100 task responses were generated for each respondent.

Once the 540 pairs of data sets were generated (each pair composed of the responses to the 100 tasks by matched unidimensional and bidimensional respondents), respondent and task parameters were estimated using BIGSTEPS (Linacre & Wright, 1997)—software that performs Rasch scaling. Pairs of task calibrations were compared using a standardized difference (Equation 2). The critical value of ±2 was chosen for rejecting the null hypothesis that the difference between a pair of calibrations was due purely to chance. Tasks whose pairs of calibrations exceeded these limits were nominated as being bidimensional, and tasks whose pairs of calibrations did not exceed these limits were nominated as being unidimensional. These nominations were compared to the "true" status (as determined during data generation) to identify the proportion of accurate nominations (for both unidimensional and bidimensional status) as well as the Type I (incorrectly nominating a task as being bidimensional) and Type II (incorrectly nominating a task as being unidimensional) error rates for task-dimension identification.

Results

Table 1 displays the overall percent of correct nominations for unidimensional and bidimensional tasks and the Type I and Type II error rates. As shown by these figures, the empirical task analytic technique resulted in an average correct nomination rate of 85% with Type I and Type II errors being about equal in frequency. The average total error rate being 15% of the nominations. That is, across all levels of the three factors investigated in this study, only 15% of the tasks, on average, were misclassified.

Table 1: Overall Decision Accuracy Rates

		Nominated	
		Unidimensional	Bidimensional
True	Unidimensional	42 %	8 %
	Bidimensional	7%	43%

Note: Average percent across all levels of each factor is shown.

To summarize the trends inherent in the factors examined in these simulated data, a three-way analysis of variance (ANOVA) was performed on the proportion of correctly nominated tasks. That is, the ANOVA examined whether the average rate of correct nomination differed among the 18 combinations (with 30 replications within each cell) of (a) the number of respondents used for task calibration (2 levels), (b) the proportion of tasks that were bidimensional (3 levels), and (c) the correlation between the two dimensions (3 levels). Because these data are proportions, it was necessary to transform them with an inverse sine transformation (Kirk, 1982). As shown in Table 2, the differences in means for all main effects and all interactions were all large enough to be statistically improbable. In addition, the effect sizes indicate that the two-way interaction between the number of respondents and the dimension

correlation was large enough to be substantively important (i.e., approached a moderate effect size of .30). This interaction accounted for 17% of the total variance.

Table 2: ANOVA Summary

Source	df	SS	MS	F*	η^2
N	1	4.14	4.14	213.33	.11
r_{12}	2	5.95	2.98	153.51	.16
p_2	2	3.24	1.62	83.58	.09
$N \times r_{12}$	2	6.31	3.16	162.79	.17
$N \times p_2$	2	3.90	1.95	100.63	.11
$r_{12} \times p_2$	4	2.07	0.52	26.73	.11
$N \times r_{12} \times p_2$	4	1.26	0.31	16.24	.03
Total	539	36.99	--	--	--

Note: N represents the number of respondents upon which tasks were calibrated, r_{12} represents the correlation between the two dimensions inherent in the task responses, and p_2 represents the proportion of tasks that were bidimensional. * p-values for all F statistics were less than .0001. The R^2 for the model was .73.

Table 3 summarizes the decision accuracy by r_{12}, the correlation between the two dimensions inherent in the tasks. As shown, the rate of incorrect nomination for unidimensional tasks decreased as the correlation between the two dimensions increased. More informative, however, is the fact that the rate of incorrect nomination of bidimensional items increased as the correlation between the two dimensions increased. This is what we would expect—as the dimensions become more similar to one another, it is more difficult to differentiate between the two dimensions. Thus, we are less able to correctly identify the tasks that are truly bidimensional. Interestingly, the overall rate of correct task nomination decreased as the correlation between the

two dimensions decreased, dropping from 90% for r_{12}=.75 to 86% for r_{12}=.50 to 80% for r_{12}=.25. This indicates that some of the extra variability added to the data by the bidimensional tasks was absorbed into the calibrations of the unidimensional tasks.

Table 3: Decision Accuracy Rate by Dimension Correlation

r_{12}=.25		Nominated: Unidimensional	Nominated: Bidimensional
True	Unidimensional	36%	14%
	Bidimensional	6%	44%
r_{12}=.50		**Unidimensional**	**Bidimensional**
True	Unidimensional	42%	8%
	Bidimensional	7%	43%
r_{12}=.75		**Unidimensional**	**Bidimensional**
True	Unidimensional	49%	1%
	Bidimensional	9%	41%

Note: Average percent across all levels of each remaining factor is shown.

Table 4 summarizes the decision accuracy by *N*, the number of respondents upon whom tasks were calibrated. These figures show that, as would be expected, the rate of bidimensional task nomination increased with larger sample sizes. This is because the more respondents upon whom the tasks are calibrated, the smaller the standard error of the task calibrations and the more likely a task is to be nominated whether due to bidimensionality or chance variation between parameter estimates. As a result, the correct nomination rate for unidimensional tasks is greater when fewer respondents are used, and the rate of correct nomination for bidimensional

tasks is greater when more respondents are used. Overall, the rate of correct nomination was better when fewer respondents were used to calibration tasks.

Table 4: Decision Accuracy Rate by Dimension Correlation

	N=500	Nominated	
		Unidimensional	Bidimensional
True	Unidimensional	48%	2%
	Bidimensional	10%	40%
	N=2000	Nominated	
		Unidimensional	Bidimensional
True	Unidimensional	36%	14%
	Bidimensional	5%	45%

Note: Average percent across all levels of each remaining factor is shown.

Table 5 summarizes the decision accuracy by p_2, the proportion of bidimensional tasks. The proportions in the diagonal cells change as a result of the number of tasks that belong to each subset. Hence, with perfect nomination accuracy, we would expect the upper left cell to equal 75% and the lower right cell to equal 25% for p_2=.25. Similarly, we would expect the upper left cell to equal 50% and the lower right cell to equal 50% for perfect nomination accuracy when p_2=.50, and we would expect the upper left cell to equal 25% and the lower right cell to equal 75% for perfect nomination accuracy when p_2=.75. It is evident from these figures that the rate of incorrect nomination for unidimensional tasks is greater when fewer bidimensional tasks exist and that the rate of incorrect nomination for bidimensional tasks is greater when more bidimensional tasks exist. Again, this result is consistent with what we would expect—the more bidimensional tasks in the data, the less likely a unidimensional task's chance variation will be erroneously

detected as bidimensionality. Overall, there was little change in the total rate of correct nomination across the proportion of tasks belonging to the second dimension.

Table 5: Decision Accuracy Rate by Proportion of Bidimensional Tasks

	p_2=.25	Nominated	
		Unidimensional	Bidimensional
True	Unidimensional	63%	12%
	Bidimensional	3%	22%
	p_2=.50	Nominated	
		Unidimensional	Bidimensional
True	Unidimensional	40%	10%
	Bidimensional	7%	43%
	p_2=.75	Nominated	
		Unidimensional	Bidimensional
True	Unidimensional	21%	4%
	Bidimensional	12%	63%

Note: Average percent across all levels of each remaining factor is shown.

Discussion

This study has demonstrated the utility of an empirical method for performing task analysis that is based on Rasch measurement theory. Overall, the method was shown to be effective as a means of identifying the specific tasks that define the differences between two groups of simulated respondents who are known to exhibit differences in their behaviors with respect to the construct being measured by the tasks. Such a method has advantages over traditional approaches to task analysis because it is much less expensive to collect and analyze task response data and because the technique does not rely on the subjective judgment of the

data analyst. In addition, the Rasch-based method has the advantage over raw score methods of not requiring every individual to respond to the same set of tasks—a constraint that would be difficult to satisfy when a large number of tasks is to be studied and funding for the study is limited.

In general, the empirical task analytic method correctly identified the dimension to which tasks belonged 85% of the time. The correlation between the dimensions and the number of respondents upon whom the tasks were calibrated seem to be the most important factors in determining the accuracy of nominations. The rate of accurate nomination increases as the correlation between the dimensions increases (although Type I and Type II error rates vary with this correlation), and more accurate nominations are made when task calibration is performed on smaller groups of respondents (500 versus 2000 in this study).

Further studies of this empirical task analysis method should focus on five issues. First, it will be important to compare the results of these analyses to those obtained from similar methods based on raw scores (e.g., factor analysis or principal component analysis). Consistency among the methods would support the validity of the technique explored in this study. Second, these simulations should be expanded to determine whether more accurate nomination procedures can be identified. For example, simulation studies performed by Smith (1994, 1996) suggest that a between item approach to evaluating fit may be less prone to Type I errors and may be less sensitive to sample size than the separate calibration approach employed in this study. Third, this empirical task analysis technique should be utilized in applied settings to determine the usefulness of the information that it provides to the development of cognitive models. Fourth, the results of this empirical task analysis technique should be compared to traditional task analysis methods to determine whether the two methods depict the nature of cognitive functioning on the tasks in question in similar ways. If the methods do not result in similar depictions, then additional work would be necessary to evaluate the validity of each approach to task analysis. Fifth, the empirical task analysis technique explored here should be extended, provided its validity can be supported, to the identification of respondents whose group membership is unknown. That is, it

would be useful to know if the task calibrations that are generated from the empirical task analysis could be used to determine to which group a respondent for whom no additional information is available belongs. Such an application would be a novel and quite useful statistical tool—not only would it be possible to empirically define the nature of differences between groups of respondents, but it would also be possible to identify the group that most closely resembles individuals who were not part of the original calibration study.

References

Ericsson, K.A., & Simon, H.A. (1982). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

DiBello, L.V., Stout, W.F., & Roussos, L.A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P.D. Nichols, S.F. Chipman, & R.L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-389). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer Nijhoff.

Kail, R.V.J., & Bisanz, J. (1982). Cognitive strategies. In C.R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 229-255). New York: Academic Press.

Kelderman, H. (1994). Objective measurement with multidimensional polytomous latent trait models. In M. Wilson (Ed.), *Objective measurement: Theory into practice* (Vol.2, pp. 235-243). Norwood, NJ: Ablex.

Kirk, R.E. (1982). *Experimental design: Procedures for the behavioral sciences*. Belmont, CA: Brooks/Cole.

Linacre, J.M., & Wright, B.D. (1997). *A user's guide to BIGSTEPS: Rasch-model computer program*. Chicago, IL: MESA.

Mumaw, R.J., Pellegrino, J.W., & Glaser, R. (1980). *Some puzzling aspects of spatial ability*. Paper presented at the annual meeting of the Psychonomic Society, St. Louis, MO.

Siegler, R.S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, *7*, 268-288.

Smith, R.M. (1994). A comparison of the power of Rasch total and between item fit statistics to detect measurement disturbances. *Educational and Psychological Measurement*, *54*, 42-55.

Smith, R.M. (1996). A comparison of Rasch separate calibration and between item methods of detecting item bias. *Educational and Psychological Measurement*, *54*, 42-55.

Tatsuoka, K.K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R.Glaser, A. Lesgold, & M.G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453-489). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wright, B.D., & Masters, G.N. (1982). *Rating scale analysis.* Chicago, IL: MESA.